

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

# Structural Characterization of Chaperone Assisted Folding and Assembly of RuBisCO

Bharathi Vasudeva Rao

aus

Kolar, Karnataka, India

2009

## **Erklärung**

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Professor Dr. F. Ulrich Hartl betreut.

## **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfen erarbeitet.

München, am .....

.....  
Bharathi Vasudeva Rao

Dissertation eingereicht am	22.01.2009
1. Gutacher	Prof. Dr. F. Ulrich Hartl
2. Gutachter	PD Dr. Konstanze F. Winklhofer
Mündliche Prüfung am	12.03. 2009



## Acknowledgements

First of all, I would like to thank Prof. Dr. F. Ulrich Hartl and Dr. Manajit Hayer-Hartl for giving me the opportunity to do my PhD research in this laboratory and for their supervision, advice, and guidance throughout the entire period of my study.

I wish to express my warm and sincere thanks to Prof. Dr. Konstanze Winklhofer for correcting my dissertation and being the co-referee of my thesis committee.

I would like to thank Dr. Andreas Bracher for all the crystallization help, Oana, Günter and Dr. Beate Rockel for all the help during the electron microscopy sessions. I would like to thank Prof. Dr. Jürgen Soll for providing us *Arabidopsis* cDNA.

I wish to extend my warmest thanks to Andrea, Silke, Elisabeth, Grampp, Emmanuel and Nadine for assisting me in many ways. I would like to thank Sandra, Kausik and all my colleagues who have helped me with my work in the Department.

I am also greatly indebted to my teachers in the past, Dr. K. M. Divakar and Late Dr. M. D. Kulashekar for their constant support and motivation.

I am immensely grateful for my husband, who is also my colleague, Karnam Vasudeva Rao who has been like a pillar of support to me. Without his love, encouragement, understanding and constant guidance, it would have been impossible for me to finish this work. My deepest gratitude is for my father, my mother and my brother for their unconditional love and support throughout my life. To them I dedicate this thesis. I would also like to thank Vasu's parents and my brother-in-law for all the support and understanding.

## Contents

<b>1 Summary</b>	<b>01</b>
<b>2 Introduction</b>	<b>03</b>
2.1 Structure of proteins	03
2.2 Protein folding	05
2.3 Disruption of the native state	07
2.4 Incorrect protein folding and disease	07
2.5 Molecular chaperones	10
2.5.1 Chaperones involved in <i>de novo</i> protein folding	11
2.5.2 Ribosome-Binding Chaperones	13
2.5.3 The Hsp70 system	14
2.5.4 Prefoldin or Gim complex	17
2.5.5 The Chaperonins	18
2.5.5.1 Group I Chaperonins	20
2.5.5.2 Chaperonins in chloroplasts and cyanobacteria	24
2.6 Photosynthesis	27
2.6.1 Light Reactions	30
2.6.2 Dark Reaction	31
2.7 Rubisco	33
2.7.1 Regulation of Rubisco	36
2.7.2 Rubisco Structure	37
2.7.3 Synthesis, Folding and assembly of Rubisco	42
2.8 RbcX	44
<b>3 Aim of the study</b>	<b>46</b>
<b>4 Materials and methods</b>	<b>48</b>
4.1 Laboratory equipment	48
4.2 Materials	48
4.2.1 Chemicals	49
4.2.2 Strains	50

## CONTENTS

---

4.2.3	Plasmids, DNA and oligonucleotides	50
4.2.4	Enzymes, proteins, peptides and antibodies	50
4.2.5	Media	51
4.2.6	Buffers and stock solutions	51
4.3	Molecular biology methods	51
4.3.1	Plasmid purification	51
4.3.2	DNA analytical methods	52
4.3.3	PCR amplification	52
4.3.4	DNA restriction digestions and ligations	53
4.3.5	Preparation and transformation of competent <i>E. coli</i> cells	54
4.3.6	Cloning strategies	55
4.4	Protein biochemical methods	56
4.4.1	Protein analytical methods	56
4.4.1.1	Determination of protein concentrations	56
4.4.1.2	SDS-PAGE	57
4.4.1.3	Native PAGE	57
4.4.1.4	Tricine-PAGE	58
4.4.1.5	Bis-Tris Native PAGE	59
4.4.1.6	Coomassie staining	59
4.4.1.7	Silver staining	60
4.4.1.8	Autoradiography	60
4.4.1.9	Western blotting	60
4.4.1.10	TCA precipitation	61
4.4.1.11	FFF-MALS	61
4.4.1.12	N-terminal sequencing of proteins	62
4.4.1.13	Sequence alignments	62
4.4.2	Protein expression and purification	62
4.4.2.1	At-ch-cpn60	62
4.4.2.2	At-ch-cpn20	63
4.4.2.3	At-ch-cpn20 <sub>N-His6</sub>	64

## CONTENTS

---

4.4.2.4	Syn6301-RbcL <sub>8</sub> S <sub>8</sub>	65
4.4.2.5	Syn6301-RbcL <sub>8</sub>	66
4.4.2.6	Syn6301-RbcS and Syn7002-RbcS <sub>FLAG</sub>	67
4.4.2.7	Syn7002-RbcX wild type and mutants	67
4.4.2.8	At-RbcX <sub>N-His6</sub>	68
4.4.2.9	At-RbcX <sub>N-His6+Ub</sub> and At-RbcX <sub>N-His6+Ub+FLAG</sub>	69
4.4.2.10	At-RbcS1A	70
4.4.3	Functional analyses	70
4.4.3.1	ATPase activity assay	70
4.4.3.2	<i>in vivo</i> co-expression of RbcL or RbcLS	71
4.4.3.3	<i>in vitro</i> translation	72
4.4.3.4	Analytical gel filtration	73
4.4.3.5	Tryptophan-fluorescence spectroscopy	74
4.4.3.6	ANS-fluorescence spectroscopy	74
4.4.3.7	Circular Dichroism Spectroscopy	75
4.4.4	Crystallography and structure analysis	75
4.4.4.1	Analytical subtilisin digestion	75
4.4.4.2	Protein crystallization	75
4.4.4.3	Structure determination	76
4.4.5	Electron microscopy	77
4.4.5.1	Preparation of At-ch-cpn60/20 complexes	77
4.4.5.2	Negative staining	77
4.4.5.3	Recording of micrographs	78
4.4.5.4	Single Particle Analysis	78
<b>5</b>	<b>Results</b>	<b>79</b>
5.1	Structural characterization of <i>A. thaliana</i> chloroplast cpn60/20	80
5.1.1	Electron microscopic analysis of At-ch-cpn60	80
5.1.2	Electron microscopic analysis of At-ch-cpn60/20	83
5.2	GroEL induced conformational changes in RbcL	87

## CONTENTS

---

5.3	Requirement of chaperonin system and RbcX for the folding and assembly of Rubisco	91
5.4	Functional similarity between GroEL/ES and ch-cpn60/20	93
5.5	Crystal structure of <i>Syn7002</i> RbcX mutants	94
5.5.1	Structure of <i>Syn7002</i> -RbcX(Q29A)	99
5.5.2	Structure of <i>Syn7002</i> -RbcX(Y17A,Y20L)	104
5.4.2.1	RbcL C-terminal peptide recognition by RbcX and the effect of Y17A,Y20L	106
5.5.3	Crystallization of <i>Syn7002</i> -RbcX(R70A)	109
5.6	Affinity of RbcX for RbcL peptide	111
5.7	Role of C-terminal peptide in the RbcX assisted RbcL <sub>8</sub> assembly	112
5.8	Characterization of <i>Arabidopsis thaliana</i> RbcX	114
5.8.1	Cloning and purification of At-RbcX	114
5.8.2	Secondary structure and predicted tertiary structure of At-RbcX	116
5.8.3	At-RbcX functions similar to <i>Syn7002</i> -RbcX	118
<b>6</b>	<b>Discussion</b>	121
6.1	Structural and functional similarity of chloroplast cpn60/20 to GroEL/ES	121
6.2	Folding and assembly of Form I Rubisco	123
6.2.1	RbcX structure and mechanism	123
6.2.2	Importance of RbcL C-terminal peptide	126
6.3	Implications for Rubisco assembly in higher plants	128
<b>7</b>	<b>References</b>	129
<b>8</b>	<b>Appendices</b>	142
8.1	Abbreviations	142
8.3	Publications	145
8.4	Curriculum vitae	146



### SUMMARY

Newly synthesized polypeptide chains must fold co- or/and post-translationally into precise three-dimensional conformation in order to become functionally active. Failure to fold correctly can result in different types of biological malfunctions. Living cells have developed molecular chaperones to prevent misfolding and aggregation of newly synthesized proteins and for the maintenance of protein homeostasis under stress condition.

Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase) catalyses the critical step in CO<sub>2</sub> fixation from the atmosphere during photosynthesis. Rubisco is inefficient and is subject to competitive inhibition by O<sub>2</sub>. This makes Rubisco an important target for possible functional improvement by protein engineering. However, such attempts are hampered by the complex folding and assembly pathway of Rubisco. The bacterial GroEL/ES system and its chloroplast homologue, the ch-cpn60/20 system, are implicated in the efficient folding of Rubisco. While the structure and function of GroEL/ES are well understood, much less is known about ch-cpn60/20. Of particular interest was cpn20, which consists of a tandem repeat of GroES. The present study, using electron microscopy, revealed that cpn60 and cpn20 form bullet shaped, asymmetric complexes, similar to GroEL/ES complexes. *in vitro* translation experiments demonstrated that ch-cpn60/20 can functionally replace GroEL/ES in the folding of cyanobacterial Rubisco large subunits. In addition to the chaperonin system, a chaperone factor, called RbcX, has been implicated in the assembly of RbcL<sub>8</sub> complexes of cyanobacterial Rubisco (Form I Rubisco).

RbcX is a dimer of ~30 kDa. In the present study, a mutational analysis of cyanobacterial RbcX from *Synechococcus* sp. PCC7002 was carried out. Mutant forms of RbcX were analysed by crystallography and demonstrated to be structurally identical to wild type RbcX. A central crevice and the peripheral corner regions of the RbcX dimer were identified as functionally critical for Rubisco assembly. The central crevice binds the

conserved C-terminal sequence, EIKFEFD, of the Rubisco large subunit. The dysfunction of the central crevice or the peripheral corner of RbcX leads to misassembly or non-specific Rubisco aggregation.

A preliminary analysis of RbcX from *Arabidopsis thaliana* revealed that higher plant RbcX and cyanobacterial RbcX share a similarity in their secondary structure and function, indicating a universal role of RbcX in Rubisco assembly. The information on Rubisco folding and assembly provided in this study may be helpful for future efforts in improving the functional properties of this enzyme.

## 2 INTRODUCTION

Proteins are the major components of living organisms and perform a wide range of essential functions in cells. The word 'protein' was first coined in 1838 to emphasize the importance of this class of molecules. The word is derived from the Greek word 'proteios' which means "of the first rank". While DNA is the information molecule, it is proteins that do the work of all cells - microbial, plant, animal. Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle.

Proteins are made of amino acids arranged in a linear chain and joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues.

### 2.1 Structure of proteins

Newly synthesized protein chains generally must fold into unique three dimensional structures in order to become functionally active. The shape into which a protein naturally folds is known as its native state. Although many proteins can fold unassisted, simply through the chemical properties of their amino acids, others require the aid of molecular chaperones to fold into their native states. Four levels of structural organization for proteins can be outlined: primary, secondary, tertiary, and quaternary. The overall conformation of a protein is the combination of these elements.

- Primary Structure refers to the linear sequence of amino acids that make up the polypeptide chain. This sequence is determined by the genetic code, the sequence of nucleotide bases in the DNA.
- Secondary structure is the ordered arrangement or conformation of amino acids in localized regions of a polypeptide or protein molecule. Hydrogen bonding plays an important role in stabilizing these folding patterns. The two main secondary structures are

the alpha helix and the anti-parallel beta-pleated sheet. A single polypeptide or protein may contain multiple secondary structures.

- The tertiary structure is the final specific geometric shape that a protein assumes. This final shape is determined by a variety of bonding interactions between the "side chains" on the amino acids. A major driving force in determining the tertiary structure of globular proteins is the hydrophobic effect. The polypeptide chain folds such that the side chains of the nonpolar amino acids are "hidden" within the structure and the side chains of the polar residues are exposed on the outer surface. Hydrogen bonding involving groups from both the peptide backbone and the side chains are important in stabilizing tertiary structure. The tertiary structure of some proteins is stabilized by disulfide bonds between cysteine residues. As a result, bonding interactions between "side chains" may cause a number of folds, bends, and loops in the protein chain. Different fragments of the same chain may become bonded together.

- Quaternary structure involves the association of two or more polypeptide chains into a multi-subunit structure. Quaternary structure is the stable association of multiple polypeptide chains resulting in an active unit. Not all proteins exhibit quaternary structure. Usually, each polypeptide within a multi-subunit protein folds more-or-less independently into a stable tertiary structure and the folded subunits associate with each other to form the final structure. Quaternary structures are stabilized mainly by noncovalent interactions. All types of noncovalent interactions: hydrogen bonding, van der Waals interactions and ionic bonding are involved in the interactions between subunits. In rare instances, disulfide bonds between cysteine residues in different polypeptide chains are involved in stabilizing quaternary structure.

In addition to these levels of structure, proteins may shift between several related structures while they perform their biological function. In the context of these functional rearrangements, these tertiary or quaternary structures are usually referred to as "conformations", and transitions between them are called 'conformational changes'. Such changes are often induced by the binding of a substrate molecule to an enzyme's active site, or the physical region of the protein that participates in chemical catalysis. In solution

all proteins undergo variation in structure through thermal vibration and the collision with other molecules (Nelson and Cox, 2005).

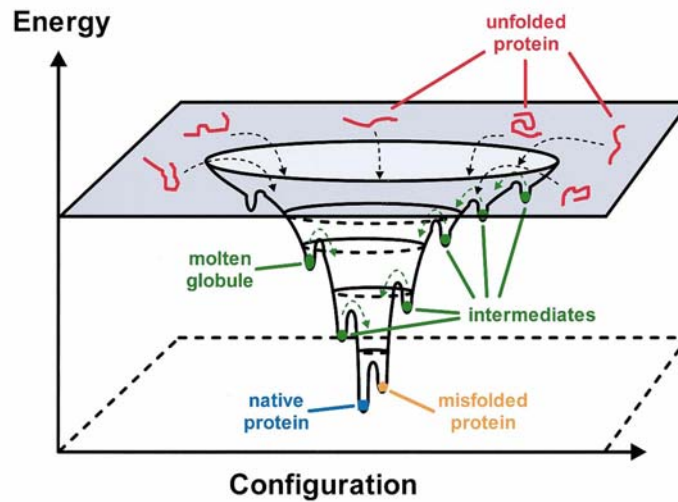
### **2.2 Protein folding**

Protein folding is a process by which a polypeptide folds into its characteristic and functional three-dimensional structure. The ability of a protein to fold reliably into a pre-determined conformation despite a near infinite number of possibilities is still poorly understood. It has long been known that the structure of a protein is determined purely by the amino acid sequence (Anfinsen, 1973) and the structure of the protein determines the function. By extension, the function of a protein depends entirely on the ability of the protein to fold rapidly and reliably to its native structure. Many proteins fold spontaneously into their native structure in aqueous solution. This folding process must satisfy two conditions - one thermodynamic, and one kinetic. The thermodynamic consideration is that the protein adopts a single, stable, folded conformation. The kinetic requirement is that the protein must fold to the native state on an appropriate timescale.

If a protein were to fold by sequentially sampling all possible conformations, it would take an enormous amount of time to do so, even if the conformations were sampled at a rapid rate; yet proteins are able to fold on a timescale of milliseconds to seconds. This suggests that only a small amount of conformational space is sampled during the folding process and this in turn implies the existence of kinetic folding pathways, as first suggested by Levinthal (Levinthal, 1968). This paradox of how proteins fold rapidly and reliably to their native conformation is known as the protein folding problem.

The major driving force in protein folding is considered to be the hydrophobic effect, which causes the formation of conformations stabilized by packing the side chains of hydrophobic amino acids into the interior of the protein. The folding process of a protein generally depends on the strength of this hydrophobic effect as well as on the stability of the protein and can be described by the energy landscape theory.

The topology of the folding energy landscape can best be described as a funnel with a rough surface (Schultz, 2000). A folding funnel is a simplified 2D or 3D representation of the very high-dimensional conformational space that the polypeptide can access during folding (Figure 1). Energetically, all possible conformations of the chain (chain entropy) is represented by the width of the funnel: the broad top of the funnel depicts the enormous number of conformations present in the soluble denatured state, such as an ensemble of starting conformations populated on rapid dilution from a urea- or guanidine-denatured state; the needle-like point at the bottom of the funnel represents the unique native structure of the protein as determined by X-ray crystallography or NMR. The separation between the top and bottom of the funnel represents other energy contributions (chain enthalpy, solvent entropy and enthalpy) to each chain conformation. The roughness of the funnel surface, and thus the formation of the native state can vary significantly between different types of proteins. Pure  $\alpha$ -helical proteins almost simultaneously undergo an enormous reduction in conformational space and hydrodynamic radius, form local segments of structure, and immediately gain some tertiary interactions.  $\alpha/\beta$  proteins initially collapse their hydrophobic cores (Li and Woodward, 1999) before assembling into ordered structures where the formation of  $\beta$ -sheets seems to be the rate-limiting step. Some folding pathways therefore include the formation of transient intermediates or more stabilized states such as the molten globule (Tsai et al., 1999).



**Figure 1. Schematic representation of the folding energy landscape of a protein molecule.**

The multiple states of the unfolded protein located at the top fall into a folding funnel consisting of an almost infinite number of local minima, each of which describes possible folding arrangements in the protein. Most of these states represent transient folding intermediates in the process of attaining the correct native fold. Some of these intermediates retain a more stable structure such as the molten globule, whereas other local minima act as folding traps irreversibly capturing the protein in a misfolded state (Schultz, 2000).

## 2.3 Disruption of the native state

In certain solutions and under some conditions proteins will not fold into their biochemically functional forms. Temperatures above (and sometimes those below) the range that cells tend to live in will cause thermally unstable proteins to unfold or "denature". High concentrations of solutes, extremes of pH, mechanical forces, and the presence of chemical denaturants can do the same. A fully denatured protein lacks both tertiary and secondary structure, and exists as a random coil. Under certain conditions some proteins can refold; however, in many cases denaturation is irreversible.

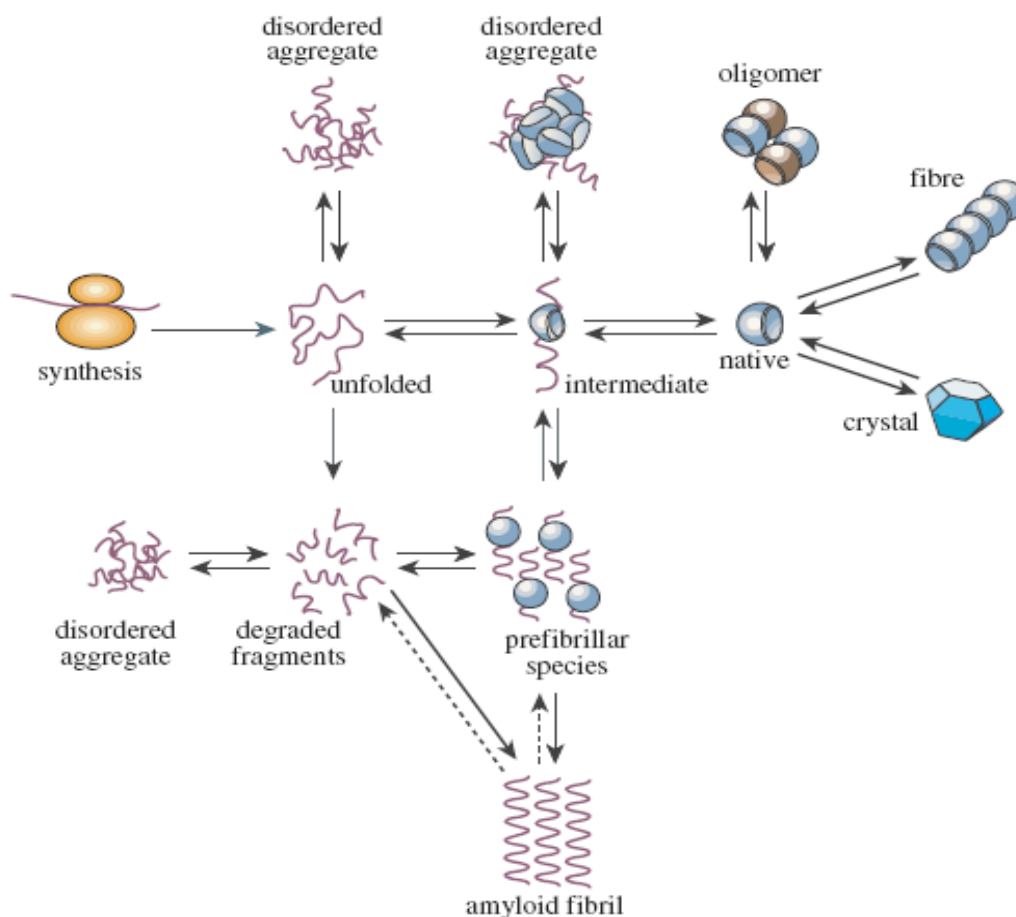
## 2.4 Incorrect protein folding and disease

Following its translation on the ribosome, the newly synthesized protein must fold into the conformation it requires in order to fulfill its biological role. This is not a trivial task because the number of theoretical interactions between each of the amino acid side

chains in a protein far exceeds the total number of protein molecules within the cell and establishing the correct interactions is vital if the protein has to fold correctly. Due to the fact that translation by ribosomes proceeds at a relatively low speed and in a vectorial manner, the elongating polypeptide chains are not available for folding at once (Frydman, 2001; Hartl and Hayer-Hartl, 2002). Also the emerging chain must face the crowded environment of the cell (300 to 400 g/liter of proteins and other macromolecules in *Escherichia coli* (Ellis, 2001) and so the chance of it making inappropriate contacts with other proteins is very high. Yet, the driving force that pushes the protein to attain its lowest free energy state (i.e. its native conformation in the majority of cases) ensures that most proteins fold spontaneously and rapidly (in the order of micro- to milliseconds) and, more often than not, folding occurs without problems. Many proteins never attain a defined conformation, and instead, remain intrinsically disordered in their biologically active state, that is, they have ill-defined secondary and tertiary structures in their native state (Ecroyd and Carver, 2008).

Multi-domain proteins refold inefficiently in contrast to the small, single-domain proteins which fold spontaneously. They could form partially folded intermediates that can aggregate (Figure 2) due to which the proteins are no longer able to enter their productive folding pathways. Owing to the high local concentration of nascent chains in polyribosomes and the added effect of macromolecular crowding, the probability of nonnative states to aggregate is increased (Hartl and Hayer-Hartl, 2002).





**Figure 2. States accessible to a protein after its biosynthesis on the ribosome**

A newly synthesized polypeptide chain can fold to a monomeric native structure from a highly disordered unfolded state often through one or more partly folded intermediates. However, it can experience other fates such as degradation or aggregation. An amyloid fibril is just one unique form of aggregate having a highly organized 'misfolded' structure. Other assemblies, including functional oligomers, macromolecular complexes and natural protein fibers, contain natively folded molecules, as do the protein crystals produced *in vitro* for X-ray diffraction studies of their structures. In living systems, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases (Vendruscolo and Dobson, 2005).

Under normal conditions, protein aggregation does occur up to a certain level in cells and this may lead to various diseases. Aggregated proteins are associated with prion-related illnesses such as Creutzfeldt-Jakob disease, bovine spongiform encephalopathy (mad cow disease), amyloid-related illnesses such as Alzheimer's

disease and familial amyloid cardiomyopathy or polyneuropathy, as well as intracytoplasmic aggregation diseases such as Huntington's and Parkinson's disease. These age onset degenerative diseases are associated with the multimerization of misfolded proteins into insoluble, extracellular aggregates and/or intracellular inclusions containing cross-beta sheet amyloid fibrils (Figure 2) (Murphy, 2002; Vendruscolo and Dobson, 2005). Misfolding and excessive degradation instead of folding and function leads to a number of proteopathy diseases such as antitrypsin-associated Emphysema, cystic fibrosis and the lysosomal storage diseases, where loss of function is the origin of the disorder.

The exposed hydrophobic patches might favour the interaction of the misfolded species with cell membranes which may lead to impairment of the function and integrity of the membranes involved, resulting in loss of regulation of the intracellular ion balance and redox status and eventually to cell death. A wide range of biological processes may be impaired by the interaction of misfolded proteins with other cellular components (Ellis, 1991).

Considering that cells are characterized by high intracellular concentrations of protein surfaces, there is a need for mechanisms to prevent incorrect interactions between these surfaces; one such mechanism is the existence of 'Molecular chaperones' (Ellis and Hartl, 1996).

### **2.5 Molecular chaperones**

Chaperones are proteins that assist the non-covalent folding/unfolding and the assembly/disassembly of other macromolecular structures, but do not occur in these structures when the latter are performing their normal biological functions.

The term 'molecular chaperone' was used first by Laskey (Laskey et al., 1978) to describe the function of nucleoplasmin to prevent the aggregation of folded histone proteins with DNA during the assembly of nucleosomes. The term was later extended by John Ellis in 1987 to describe proteins that mediate the post-translational assembly of

protein complexes (Ellis, 1991). Later it was realized that similar proteins mediate this process in both prokaryotes and eukaryotes.

Many chaperones are expressed constitutively and their level increases under stress conditions (Gething and Sambrook, 1992) and hence chaperones are termed as stress proteins or heat-shock proteins (Hsps). In general, chaperones recognize hydrophobic residues and/or unstructured backbone regions in their substrates. Chaperone binding not only prevents intermolecular aggregation by shielding the interactive surfaces of non-native polypeptides, but also prevents or reverses intramolecular misfolding (Hartl and Hayer-Hartl, 2002).

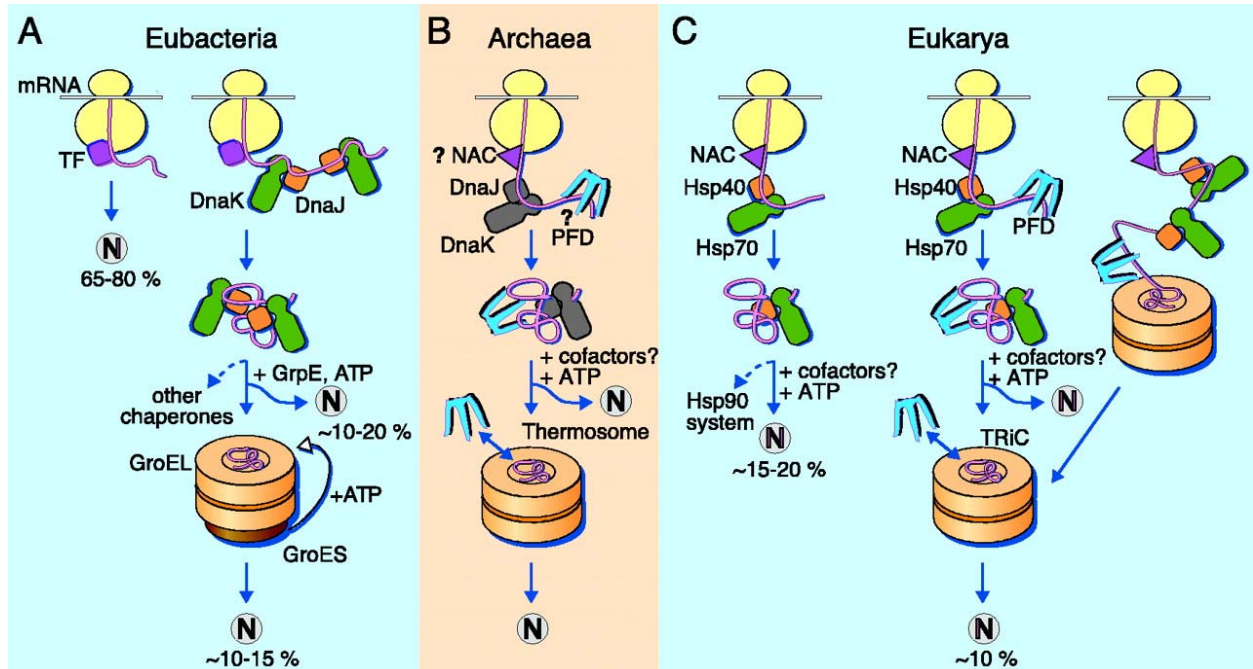
Molecular chaperones are found in all compartments of a cell where folding or, conformational rearrangements of proteins occur. Based on their molecular weight, molecular chaperones are divided into several classes or families. A cell may express multiple members of the same chaperone family. There exists significant amount of sequence homology among the members of the same class of molecular chaperones and they might be structurally and functionally related, while chaperones from different families are structurally unrelated (Walter and Buchner, 2002). Major classes of chaperones are Hsp100 (ClpA/B/X), Hsp90 (HtpG), Hsp70 (DnaK), the chaperonin-Hsp60 (GroEL), and the small Hsps (IbpA/B). These chaperones exhibit housekeeping functions under physiological conditions and damage-control functions under stress conditions (Ben-Zvi and Goloubinoff, 2001).

### **2.5.1 Chaperones involved in *de novo* protein folding**

In the cytosol, *de novo* protein folding is accomplished by two distinct sets of chaperones. Trigger factor and the Hsp70s hold the nascent and newly synthesized chains in a state competent for folding upon release into the medium. In contrast, the large, cylindrical chaperonin complexes provide physically defined compartments inside which a complete protein or a protein domain can fold while being sequestered from the cytosol. These two classes of chaperones are conserved in all three domains of life (Figure 3) and can cooperate in a topologically and timely ordered manner (Hartl and

Hayer-Hartl, 2002). Some of these chaperones, including trigger factor and specialized Hsp70 proteins, bind directly to the ribosome near the polypeptide exit site and are positioned to interact generally with nascent chains.

Most of the small proteins fold rapidly, without further assistance upon completion of synthesis and release from this first set of components. Longer chains interact subsequently with members of a second class of nascent chain-binding chaperones, including Hsp70s and prefoldin (Figure 3), which do not associate directly with the ribosome. In addition to stabilizing elongating chains, these chaperones also assist in co- or posttranslational folding, or facilitate chain transfer to downstream chaperones. A subset of slow-folding and aggregation-sensitive proteins (10 to 15% of total) interacts with a chaperonin for folding in both prokaryotes and eukaryotes. Many eukaryotic kinases and other signal-transduction proteins use an additional chaperone pathway from Hsp70 to Hsp90 (Figure 3), a specialized ATP-dependent chaperone that cooperates with ancillary factors in protein folding and regulation (Hartl and Hayer-Hartl, 2002).



**Figure 3. Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol of Eubacteria, Archaea and Eukarya (Hartl and Hayer-Hartl, 2002).**

A. In Eubacteria, nascent chains probably interact generally with trigger factor (TF), and most small proteins (~65 to 80% of total) fold rapidly upon synthesis without further assistance. Longer chains (10 to 20% of total) interact subsequently with DnaK and DnaJ and fold upon one or several cycles of ATP-dependent binding and release. About 10 to 15% of chains transit the chaperonin system--GroEL and GroES--for folding. GroEL does not bind to nascent chains and is thus receive an appreciable fraction of its substrates after their interaction with DnaK.

(B) Archaea. PFD, prefoldin; NAC, nascent chain-associated complex. Only some archaeal species contain DnaK/DnaJ. The existence of a ribosome-bound NAC homolog, as well as the interaction of PFD with nascent chains, has not yet been confirmed experimentally.

(C) In Eukarya, NAC probably interacts generally with nascent chains. The majority of small chains may fold upon ribosome release without further assistance. About 15 to 20% of chains reach their native states in a reaction assisted by Hsp70 and Hsp40, and a fraction of these must be transferred to Hsp90 for folding. About 10% of chains are co- or posttranslationally passed on to the chaperonin TRiC in a reaction mediated by PFD.

### 2.5.2 Ribosome-Binding Chaperones

In *E. coli*, Trigger factor (TF) is the first chaperone to meet nascent polypeptides (Figure 3A) as they emerge from ribosomes. It is 48 kDa in size and binds to ribosomes at a 1:1 stoichiometry (Hesterkamp et al., 1996). TF exhibits peptidyl-prolyl cis/trans isomerase (PPIase) activity *in vitro*; however, independent of proline residues, TF recognizes the target polypeptides enriched in hydrophobic (aromatic) amino acids (Patzelt et al., 2001). TF has an overlapping chaperone function with the main bacterial Hsp70 system, DnaK and DnaJ, in stabilizing nascent chains in a state competent for subsequent folding. *E. coli* cells lacking TF ( $\Delta$ tig) or DnaK ( $\Delta$ dnak) exhibit no apparent folding defects at 37°C; however, deletion of dnaK in a  $\Delta$ tig strain is lethal (Genevaux et al., 2004). TF docks onto the ribosomal L23 protein and also contacts L29, both of which are near the polypeptide exit site of the large ribosome subunit. Because TF is an ATP-independent chaperone, it does not actively assist folding through nucleotide-regulated cycles of polypeptide binding and release. TF has been shown to increase the folding efficiency of certain multidomain proteins concomitant with delaying their folding relative to translation (Agashe et al., 2004). TF is thought to function by scanning a nascent chain as it is extruded from the exit site and shielding hydrophobic stretches to keep them soluble. Active folding of the nascent and newly synthesized polypeptides can then be

mediated by ATP-dependent chaperones, such as DnaK or GroEL (Hartl and Hayer-Hartl, 2002).

In the eukaryotic cytosol, with TF being absent, NAC (nascent chain-associated complex) functions as the ribosome associated chaperone. It is a heterodimeric complex of  $\alpha$  (33 kDa) and  $\beta$  (22 kDa) subunits (Wiedmann et al., 1994). NAC binds to short nascent chains and dissociates upon chain release from the ribosome (Beatrix et al., 2000; Hartl, 1996).

Yeast and other fungi have cytosolic Hsp70 homologs that are specialized in nascent chain binding. The Ssb1 and Ssb2 proteins in the yeast *Saccharomyces cerevisiae* interact with the ribosome and with short nascent chains. This function of the Ssb proteins appears to be mediated by another Hsp70, Ssz1, which forms a stable ribosome-associated complex with zuotin (Gautschi et al., 2001), the Hsp40 partner of Ssb1 and Ssb2 (Bukau et al., 2000). RAC and the Ssb proteins act in concert in stabilizing nascent chains (Hartl and Hayer-Hartl, 2002).

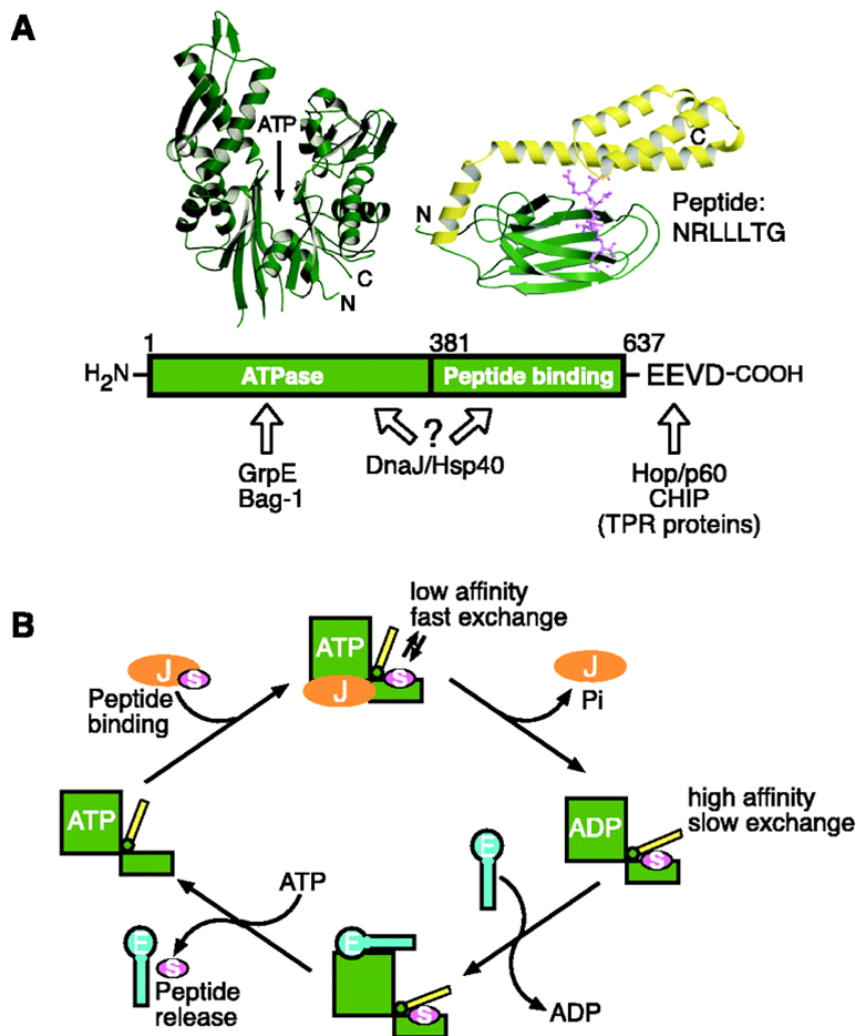
### 2.5.3 The Hsp70 system

Hsp70s are a highly conserved family of proteins, distributed ubiquitously in all prokaryotes and in cellular compartments of eukaryotic organisms. There are multiple homologs of Hsp70 with distinct cellular functions in some compartments. The yeast *S. cerevisiae* has four non-ribosome-binding Hsp70 proteins in the cytosol, namely, Ssa1 to Ssa4 and three ribosome-associated Hsp70s, called Ssb1, Ssb2, and Pdr13 (or Ssz1). Only Ssa-type function is essential for viability and Ssb activity cannot substitute for Ssa activity. Higher eukaryotes have both constitutively expressed Hsp70 homologs (Hsc70) and stress-inducible forms (Hsp70) in their cytosol. In an ATP-dependent manner, Hsp70s, assisted by Hsp40 (DnaJ) cochaperones, function by binding and releasing the extended polypeptide segments exposed by non-native states of the proteins (Hartl and Hayer-Hartl, 2002).

Hsp70 assists a wide range of folding processes, including the folding and assembly of newly synthesized proteins, refolding of misfolded and aggregated proteins, membrane translocation of organellar and secretory proteins, and control of the activity of regulatory proteins (Bukau et al., 2000; Hartl and Hayer-Hartl, 2002; Young et al., 2003). Hsp70s have thus housekeeping functions in the cell in which they are built-in components of folding and signal transduction pathways, and quality control functions in which they proofread the structure of proteins and repair misfolded conformers (Mayer and Bukau, 2005).

The Hsp70 homolog of *E. coli* (DnaK), its Hsp40 cochaperone (DnaJ) and the nucleotide exchange factor (GrpE) are very well characterized structurally and mechanistically. DnaK consists of an NH<sub>2</sub>-terminal ATPase domain of ~44-kDa and a COOH-terminal peptide-binding domain of ~27-kDa (Figure 4A). The latter is divided into a  $\beta$ -sandwich subdomain with a peptide-binding cleft and an  $\alpha$ -helical latch-like segment. The exposed hydrophobic amino acid side chains in conjunction with an accessible polypeptide backbone, mostly found in the nascent chains, are recognized by DnaK. The  $\alpha$ -helical latch over the peptide-binding cleft is in an open conformation in the ATP-bound state of DnaK (Figure 4B). It is in this state that rapid peptide binding occurs. Upon hydrolysis of ATP to ADP, the latch closes holding the peptide in a stable state. DnaJ and GrpE regulate the cycling of DnaK between its different nucleotide-bound states. The 41-kDa DnaJ protein can bind to unfolded polypeptides and prevent their aggregation (Langer et al., 1992a; Rudiger et al., 2001). DnaJ binds to DnaK and stimulates its ATPase activity, generating the ADP-bound state of DnaK, which interacts stably with the polypeptide substrate. The 23-kDa GrpE protein acts as a nucleotide exchange factor; it binds to the ATPase domain of DnaK and, by distorting the nucleotide binding pocket, induces release of bound ADP. Finally, rebinding of ATP triggers dissociation of the DnaK-substrate complex (Hartl and Hayer-Hartl, 2002).





**Figure 4. Structure and reaction cycle of DnaK system.**

(A) Structures of the ATPase domain and the peptide-binding domain of DnaK. The  $\alpha$ -helical latch of the peptide binding domain is shown in yellow and an extended peptide substrate in pink. ATP indicates the position of the nucleotide binding site. The interaction of prokaryotic and eukaryotic cofactors with Hsp70 is shown schematically. Only the Hsp70 proteins of the eukaryotic cytosol have the COOH-terminal sequence EEVD that is involved in binding of tetratricopeptide repeat (TPR) cofactors.

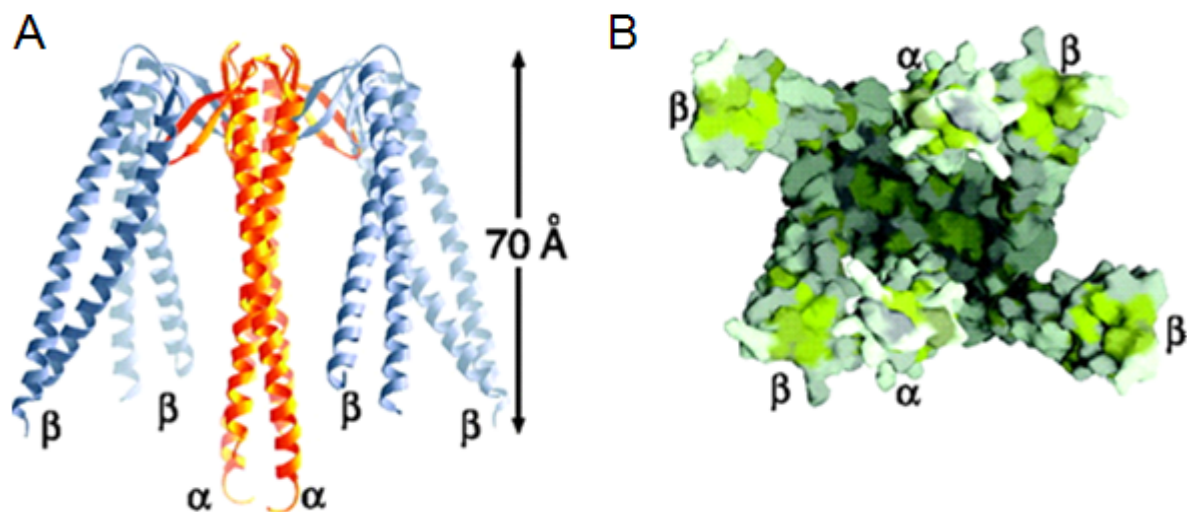
(B) Simplified reaction cycle of the DnaK system with DnaK colored as in (A). J, DnaJ; E, GrpE; S, substrate peptide. The cycle starts with the association of non native substrate proteins with either DnaJ (J) or DnaK (K) in the ATP-bound open conformation. DnaJ and substrate protein then stimulate the ATP hydrolysis of DnaK, resulting in the closure of its substrate binding pocket. The interaction of GrpE (E) with DnaK promotes the exchange of bound ADP for ATP. This results in the opening of the substrate binding cleft and the exchange of substrate proteins. The released protein can then either fold towards the native state or rebind to DnaJ or DnaK (Hartl and Hayer-Hartl, 2002).



Hsp70 also facilitates the posttranslational folding of multidomain proteins through cycles of binding and release. On release from Hsp70, an unfolded chain is free to partition to its native state. Slow-folding intermediates may rebind to Hsp70 and thus escape aggregation.

### **2.5.4 Prefoldin or Gim complex**

In eukaryotic cells and archaea, the Gim complex (GimC, genes involved in microtubule biogenesis) or prefoldin, respectively, acts in an Hsp70-like manner in stabilizing nascent chains. Prefoldin (PFD) is a jelly fish like 90-kDa complex of two  $\alpha$  and four  $\beta$  subunits (Figure 5). The six  $\alpha$ -helical coiled-coil tentacles arising from the  $\beta$ -barrel body expose hydrophobic amino acid residues for the binding of nonnative protein (Figure 5A and 5B). Substrate binding and release by PFD is ATP independent. *in vitro*, mammalian and archaeal PFD can stabilize nonnative proteins for subsequent transfer to a chaperonin (Vainberg et al., 1998). In eukaryotes, PFD in cooperation with the chaperonin TRiC assists the folding of actin and tubulin (Siegers et al., 1999). In yeast, combined deletion of the Ssb-class Hsp70s and of PFD resulted in a pronounced synthetic growth defect similar to the synthetically lethal phenotype of the TF and DnaK deletions in *E. coli* (Deuerling et al., 1999). This implicated that PFD in the archaeal cytosol may have a DnaK or TF-like role.



**Figure 5. Structure of archaean Prefoldin.**

A. Side view and dimension of the structure of archaean PFD with the two  $\alpha$  subunits shown in gold and the four  $\beta$  subunits in gray.

B. Bottom view of the PFD complex showing the central space enclosed by the six coiled-coil segments (Hartl and Hayer-Hartl, 2002).

### 2.5.5 The Chaperonins (Ring-Shaped Chaperones)

The chaperonins are a group of essential proteins, highly related by sequence, identified in all three kingdoms of life. They form large (800-1000 kDa) oligomeric cylindrical complexes consisting of two stacked rings, each enclosing a central cavity to which unfolded polypeptides bind and reach the folded state. Chaperonins differ substantially from Hsp70 in architecture, as well as in their mechanism. Similar to Hsp70, ATP binding and hydrolysis induces conformational changes in them (Hartl and Hayer-Hartl, 2002). This drives the substrate binding and release cycle. Nonnative substrate protein is first captured through hydrophobic contacts with multiple chaperonin subunits and is then displaced into the central cavity where it folds, protected from aggregating with other nonnative proteins.

Based on their sequence homology, they can be classified into two groups.

Group I chaperonins (or Hsp60s) consist of members from eubacteria (GroEL), mitochondria (Hsp60), and chloroplasts (ch-cpn60). They cooperate with ring-shaped cofactor/cofactors of the Hsp10 family (GroES), which form a lid on the cage. *E.coli* GroEL and its cofactor GroES are the best studied examples of the Group I chaperonins.

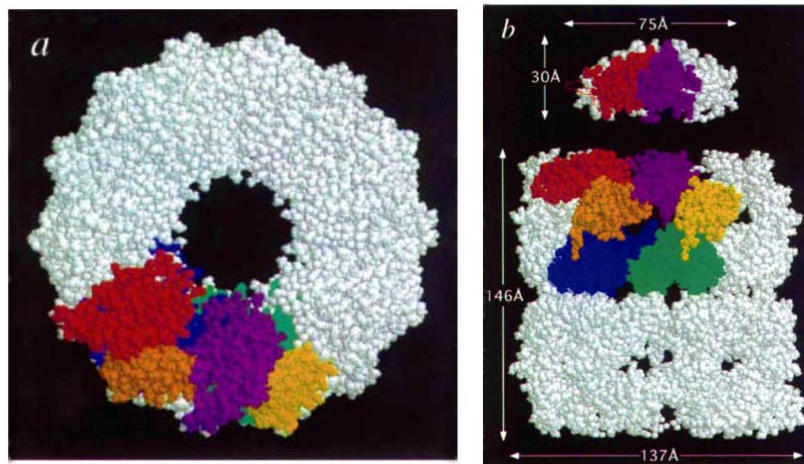
GroEL is a barrel shaped complex with fourteen identical subunits (60 kDa each) assembled into two seven-membered rings, thereby forming two separate cavities. The co-chaperone GroES has a dome-shaped ring-structure and consists of seven subunits of 10 kDa each (Hartl and Hayer-Hartl, 2002).

Group II chaperonins consist of members from archaea (thermosome) and the eukaryotic cytosol (TRiC or CCT; TCP-1 ring complex or chaperonin-containing TCP-1, respectively, where TCP-1 is tailless complex polypeptide-1) (Figure 3C). They also form a cage-like structure and are hetero-oligomeric complexes. TRiC is a ring shaped complex consisting of eight different, yet homologous, subunits (between 50 and 60 kDa) per ring (Spiess et al., 2004; Valpuesta et al., 2002). The archaeal thermosomes have two or three different subunits per complex (arranged in eight- or nine-fold symmetrical rings respectively). The Group II chaperonins lack a co-chaperone and their apical domain has an  $\alpha$ -helical insertion that protrudes from the ring opening and functions as a built-in lid of the central cavity (Gutsche et al., 1999; Leroux and Hartl, 2000; Meyer et al., 2003). The eukaryotic chaperonin TRiC can assist the folding of proteins such as actin and tubulin that cannot be folded by the GroEL/ES chaperone system (Hartl and Hayer-Hartl, 2002). ATP binding induces encapsulation of the protein by the apical-domain protrusions and initiates folding through its built-in lid mechanism. TRiC transiently interacts with ~9-15% of newly synthesized proteins ranging from 30 to 120 kDa in size. TRiC may act cotranslationally in the folding of discrete domains of proteins that are too large to be encapsulated as a whole (Frydman, 2001; Hartl and Hayer-Hartl, 2002).

### 2.5.5.1 Group I Chaperonins

The group I chaperonins, especially the GroEL system (Figure 3A) is perhaps the best characterized of all chaperones. GroEL and GroES (Figure 6) were first identified as proteins essential for bacteriophages  $\lambda$  and T5 head morphogenesis and for bacteriophage T4 tail assembly in *E. coli*. The *groE* genes were originally named so because mutations in them blocked  $\lambda$  growth and the first compensatory mutations characterized were mapped in the  $\lambda$ E gene (hence, *groE*); L and S stand for large and small gene products (Georgopoulos, 1992).

GroEL is a tetradecameric complex consisting of identical 57 kDa subunits. Electron microscopic analysis revealed that GroEL has a double-toroidal structure with approximate dimension of  $\sim 140$  Å in diameter and 150 Å in height. The central cavity of the cylinder, the site for polypeptide binding, is  $\sim 50$  Å wide (Hartl, 1996). The crystal structure of GroEL showed that each 57kDa subunit is composed of three domains: the equatorial domain containing the ATP binding pockets, the apical domains containing a patch of hydrophobic amino acids (that face the interior of the cavity and bind the unfolded substrate polypeptide through hydrophobic contacts) and an intermediate, hinge-like domain that connects the other two domains (Figure 6b). The co-chaperone GroES (Figure 6b) is a dome-shaped ring-structure with a diameter of 75 Å and consists of seven subunits. GroES has a stretch of 16 amino acids forming a mobile loop and these mediate its binding to GroEL. GroEL interacts with GroES in an adenosine nucleotide-dependent fashion (Chandrasekhar, 1986; Tilly et al., 1981), usually forming a bullet-like structure in electron micrographs (Ishii et al., 1992; Langer et al., 1992b; Saibil et al., 1991). In the presence of ATP, high magnesium concentration, and a high pH, EM studies show a significant increase in the proportion of football-like structures, with GroES binding at both ends (Azem et al., 1995; Azem et al., 1994; Schmidt, 1994). There is negative cooperativity between the two GroEL rings; they do not occur in the same nucleotide- bound state.



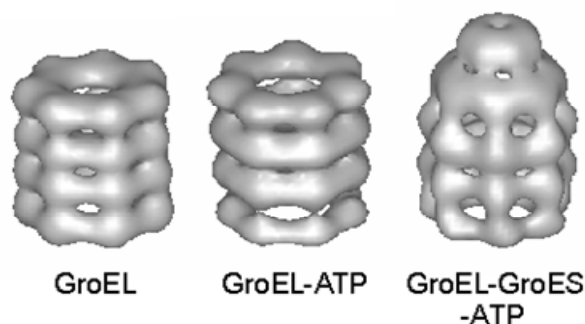
**Figure 6. Structure of the GroEL-GroES chaperone system.**

a, b, Space-filling representations showing a top and side view, respectively, of the crystal structure of GroEL alone and GroEL in complex with GroES. Two adjacent subunits are coloured with the apical domains in red and purple, the intermediate domains in orange and yellow, and the equatorial domains in blue and green, respectively (Hartl, 1996).

The chaperonin reaction (Figure 7) begins by the binding of substrate polypeptide to the free end (i.e., the *trans* ring) of a GroEL/ES complex. This is followed by the binding of seven ATP molecules and GroES, resulting in the displacement of substrate into a GroES capped cavity and causing the dissociation of the seven ADP molecules and GroES from the former *cis* complex. The apical domains, upon binding to GroES undergo a massive rotation and upward movement, resulting in an enlargement of the cavity and a shift in its surface properties from hydrophobic to hydrophilic. Non-native proteins up to ~60 kDa can be encapsulated and are free to fold in the resulting GroEL-GroES cage. Folding is allowed to proceed for ~10 s, timed by the hydrolysis of the seven ATP molecules in the *cis* ring. Upon completion of hydrolysis, binding of seven ATP molecules to the *trans* ring triggers the opening of the cage. Both folded and nonnative protein exit at this point, but folding intermediates that still expose extensive hydrophobic surfaces are rapidly recaptured and folding cycles are repeated until the protein reaches its native state (Hartl and Hayer-Hartl, 2002). Oligomeric assembly occurs in solution after subunit folding inside the cage. In addition to preventing aggregation during folding,

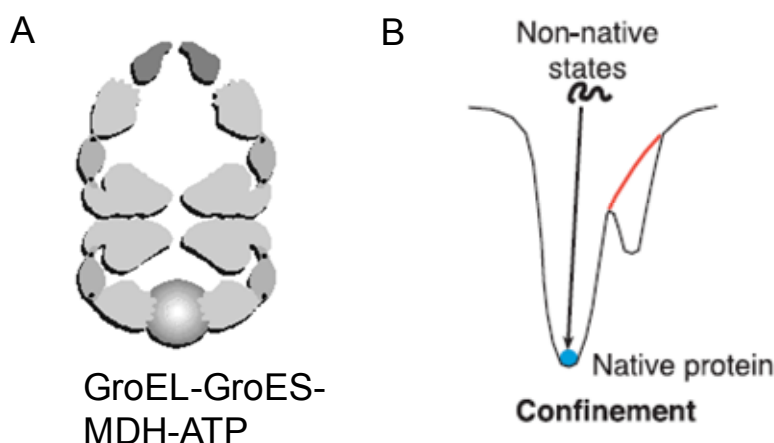


them from aggregating. In contrast, class III substrates depend stringently upon the GroEL/ES system to refold at 37°C; this class does not use the DnaK/J system to refold, even though the latter chaperones bind to them (Ellis, 2005; Kerner et al., 2005).



**Figure 8.** 3D reconstructions of GroEL, GroEL-ATP and GroEL-GroES-ATP from cryo-EM. The GroES ring is seen as a disk above the GroEL ([www.cryst.bbk.ac.uk](http://www.cryst.bbk.ac.uk)).

The obligate GroEL substrates include at least 13 essential proteins, implicating the indispensability of the chaperonin system for *E. coli* viability. The limited set of class III proteins (less than 5% of total) define the core cytosolic proteins with an obligate dependence on a specific chaperone mechanism, suggesting a high degree of folding robustness of the *E. coli* proteome as a result of an extensive functional redundancy among chaperone classes. Proteins with the  $(\beta\alpha)_8$  TIM-barrel fold are highly enriched among class III substrates (Kerner et al., 2005).



**Figure 9.** The major functional sites in GroEL and the mechanism of accelerated folding.



A. Schematic diagram of the subunit arrangement in a hypothetical slice through the GroEL oligomer. The folding complex GroEL-MDH-GroES-ATP, with the subunit domains, GroES (dark gray) and substrate (shaded) densities are based on cryo EM observations.

B. Mechanisms of accelerated folding; Confinement of nonnative protein in the narrow, hydrophilic environment of the GroEL-GroES cage is suggested to result in a smoothing of the energy landscape such that formation of certain trapped intermediates is avoided (Hartl and Hayer-Hartl, 2002).

### 2.5.5.2 Chaperonins in Chloroplast and cyanobacteria

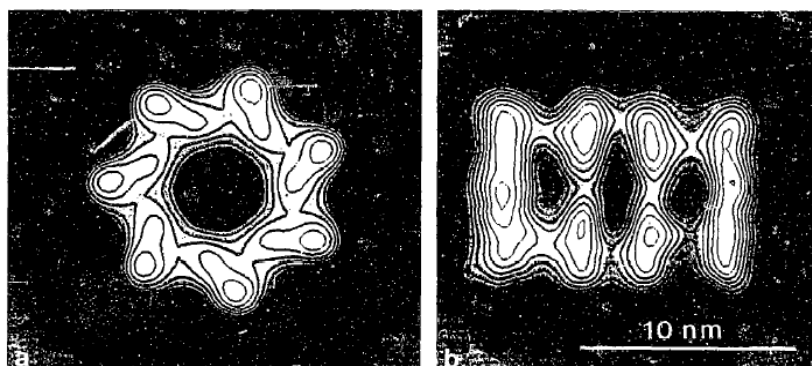
Chloroplasts contain a full set of chaperones belonging to the Hsp100, Hsp70, and Hsp60 protein families.

Nuclear-encoded group I chaperonins are found within the plastids and mitochondria of plant cells, and it was based upon studies of the chloroplast cpn60 (or Rubisco large-subunit-binding protein) that the molecular chaperone concept was revived and extended (Hemmingsen et al., 1988). The chloroplast chaperonin 60 (ch-cpn60) was initially identified as an abundant oligomeric protein that transiently binds the nascent large subunits of ribulose-1,5-bisphosphate carboxylase (Rubisco) prior to their assembly into the Rubisco holoenzyme.

The ch-cpn60 is synthesized as a nuclear-encoded precursor that is subsequently imported into chloroplasts. The protein is constitutively expressed, although its levels increase slightly during heat-shock. Similar to GroEL, the native ch-cpn60 is a cylindrical 14-mer comprised of two stacked rings with sevenfold symmetry (Figure 10) and exhibits a weak intrinsic ATPase activity (Boston et al., 1996). Ch-cpn60 appears to be hetero-oligomeric, consisting of two divergent subunits,  $\alpha$  (61 kDa) and  $\beta$  (60 kDa) (Hemmingsen and Ellis, 1986; Musgrove et al., 1987), that are no more similar to each other than they are to GroEL (Martel et al., 1990). The  $\alpha$  and  $\beta$  subunits of *A. thaliana* cpn60 are approximately 51% identical to each other and approximately 45% identical to the mitochondrial cpn60 protein. These intersubunit identities are similar to those found between prokaryotic cpn60 homologues and any of the eukaryotic subunits: the *E. coli* GroEL is 48%, 52%, and 57% identical to the  $\alpha$ ,  $\beta$ , and mitochondrial cpn60 proteins of *A. thaliana*, respectively. It remains unclear if there are distinct  $\alpha$  and  $\beta$  ring structures *in vivo* or if there are structures composed of varying proportions of  $\alpha$  and  $\beta$  subunits. Purified  $\beta$



subunits can self-assemble into functional 14-mers that are active in protein folding (Dickson et al., 2000). The  $\beta$  assembly reaction requires adenine nucleotides, is highly dependent on protein concentration, and is potentiated by GroES homologues. In contrast, purified  $\alpha$  subunits only incorporate into 14-mers in the presence of  $\beta$  subunits and this reaction also depends on adenine nucleotides (Boston et al., 1996). Analysis of the *A. thaliana* genome has revealed four homologues of the  $\beta$  subunit and two homologues of the  $\alpha$  subunit that are targeted to chloroplasts (Hill and Hemmingsen, 2001).

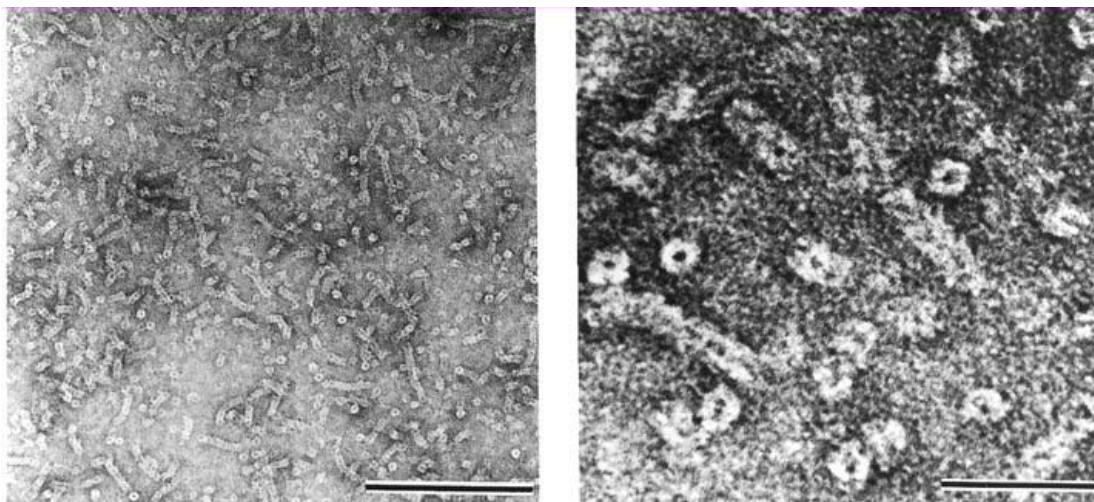


**Figure 10. Molecular projections of chloroplast cpn60 of pea.**

Correlation averages of top view and side view of ch-cpn60 isolated from pea (Tsuprun et al., 1991).

Plastids contain two types of nuclear encoded co-chaperones or co-factors; cpn20 (sometimes designated as cpn21) and cpn10. The chloroplast cpn20 (21 kDa) (Figure 11) is a functional homologue of the mitochondrial cpn10. cpn20 comprises two GroES-like domains fused head-to-tail through a short intervening linker and has twice the size of GroES and mitochondrial cpn10s (Hartl, 1996). Both halves of the molecule are highly conserved at a number of residues that are thought to be important for cpn10 function (Bertsch et al., 1992), and each contains a mobile loop region analogous to that of GroES. The two GroES-like domains show 46% sequence identity to each other (Hirohashi et al., 1999). Tetramers of cpn20 were detected when a cDNA corresponding to *A. thaliana* cpn20 was expressed in *E. coli* (Hirohashi et al., 1999; Koumoto et al.,

1999). It has been observed that *A. thaliana* cpn20 is a calmodulin-binding protein and that the calcium-calmodulin messenger system may be involved in regulating Rubisco assembly in the chloroplast (Yang and Poovaiah, 2000). The ch-cpn20 was functional *in vitro* and could assist both GroEL and ch-cpn60, purified from chloroplasts, in the refolding of *R. rubrum* Rubisco and mammalian mitochondrial malate dehydrogenase (MDH) (Baneyx et al., 1995; Bertsch and Soll, 1995). Similar to GroES, it was able to partially inhibit the ATPase activity of GroEL and ch-cpn60. Each domain was able to support bacteriophage growth when expressed individually in *E. coli*, although to a lesser extent than the double-domain cpn20.



**Figure 11: Electron micrographs of recombinant chloroplast cpn20 of spinach.**

The protein fixed with 1% glutaraldehyde and negatively stained with 1% uranyl acetate is shown at two magnifications 100 nm and 20 nm (Baneyx et al., 1995).

Other than ch-cpn20, *A. thaliana* contains cpn10 co-chaperones that are directed to the chloroplast (Hill and Hemmingsen, 2001). ch-cpn10 has been shown to be almost as efficient as GroES in assisting GroEL-mediated protein refolding. Like the bacterial and mitochondrial cpn10s, ch-cpn10 assembles into heptamers in solution (Sharkia et al., 2003). It has been reported that the mRNA for chl-cpn10 is present in the leaves and stems, not in the roots while mRNA for ch-cpn20 is abundant in leaves and was also present in roots and stems (Koumoto et al., 2001; Koumoto et al., 1999).

## 2.6 Photosynthesis

Photosynthesis is the fundamental process by which plants, algae and photosynthetic bacteria (like cyanobacteria) utilize the energy of sunlight to convert carbon dioxide and water into carbohydrates, with the green pigment chlorophyll acting as the energy converter. In plants, algae and certain types of bacteria, the photosynthetic process results in the release of molecular oxygen and the removal of carbon dioxide from the atmosphere that is used to synthesize carbohydrates (oxygenic photosynthesis). Other types of bacteria use light energy to create organic compounds but do not produce oxygen (anoxygenic photosynthesis). Photosynthesis is one of the most important biochemical pathways providing the energy and reduced carbon required for the survival of virtually all life on our planet, as well as the molecular oxygen necessary for the survival of oxygen consuming organisms.

Photosynthesis uses light energy and carbon dioxide to make triose phosphates (G3P). G3P is generally considered the first end-product of photosynthesis. It can be used as a source of metabolic energy, or combined and rearranged to form monosaccharide or disaccharide sugars, such as glucose or sucrose, respectively, which can be transported to other cells, stored as insoluble polysaccharides such as starch, or converted to structural carbohydrates, such as cellulose or glucans.

A simplified equation for photosynthesis is:

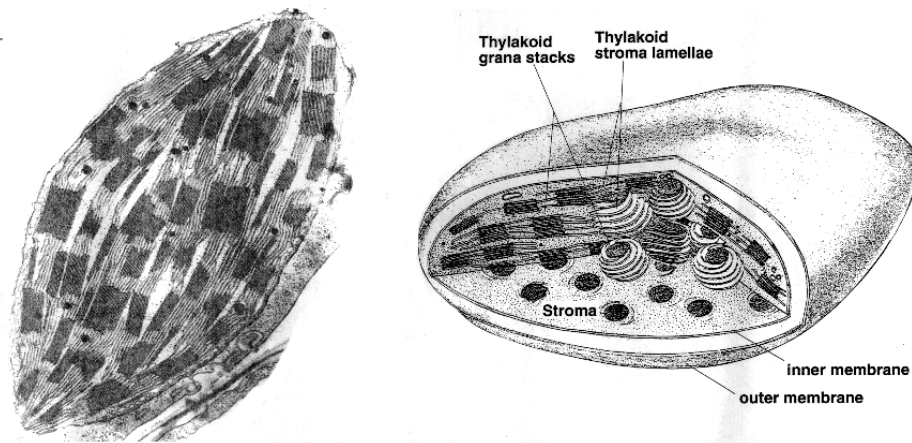


Photosynthesis occurs in two stages. In the first stage, light-dependent reactions or photosynthetic reactions capture the energy of light and use it to make high-energy molecules. During the second stage, the light-independent reactions (also called the Calvin-Benson Cycle, and formerly known as the Dark Reactions) use the high-energy molecules to capture and chemically reduce carbon dioxide (carbon fixation) to make the precursors of carbohydrates.

The photosynthetic process in plants and algae occurs in small organelles known as chloroplasts that are located inside cells. The more primitive photosynthetic organisms, like oxygenic cyanobacteria, prochlorophytes and anoxygenic photosynthetic bacteria, lack organelles.

Chloroplasts are disk-shaped structures ranging from 5 to 10 micrometers in length (Figure 12). Like mitochondria, chloroplasts are surrounded by an inner and an outer membrane. The inner membrane encloses a fluid-filled region called the stroma that contains enzymes for the light-independent reactions of photosynthesis. Infolding of this inner membrane forms interconnected stacks of disk-like sacs called thylakoids, often arranged in stacks called grana (Figure 12). The thylakoid membrane, that encloses a fluid-filled thylakoid interior space, contains chlorophyll and other photosynthetic pigments (carotenoids, phycocyanins) as well as electron transport chains. Energy trapped from sunlight by chlorophyll is used to excite electrons in order to produce ATP by photophosphorylation. The light-dependent reactions that trap light energy and produce the ATP and NADPH needed for photosynthesis occur in the thylakoids. The light-independent reactions of photosynthesis use this ATP and NADPH to produce carbohydrates from carbon dioxide and water, a series of reactions that occur in the stroma of the chloroplast. The outer membrane of the chloroplast encloses the intermembrane space between the inner and outer chloroplast membranes. Chloroplasts replicate giving rise to new chloroplasts as they grow and divide. They also have their own DNA and ribosomes.

Photosynthetic bacteria do not have chloroplasts (or any membrane-bound organelles). Instead, photosynthesis takes place directly within the cell. Cyanobacteria contain thylakoid membranes very similar to those in chloroplasts and are the only prokaryotes that perform oxygen-generating photosynthesis.



**Figure 12 Structure of a chloroplast**

Chloroplasts are surrounded by an inner and an outer membrane. The inner membrane encloses a fluid-filled region called the stroma that contains enzymes for the light-independent reactions of photosynthesis. Infolding of this inner membrane forms interconnected stacks of disk-like sacs called thylakoids, often arranged in stacks called grana. The thylakoid membrane, that encloses a fluid-filled thylakoid interior space, contains chlorophyll and other photosynthetic pigments as well as electron transport chains. The light-dependent reactions of photosynthesis occur in the thylakoids. The outer membrane of the chloroplast encloses the intermembrane space between the inner and outer chloroplast membranes (Ort, 1994).

The other photosynthetic bacteria have a variety of different pigments, called bacteriochlorophylls, and use electron donors different from water and thus do not produce oxygen. Some bacteria, such as *Chromatium*, oxidize hydrogen sulfide instead of water for photosynthesis, producing sulfur as waste. Few other photosynthetic bacteria oxidize ferrous iron to ferric iron, others nitrite to nitrate, and still others use arsenites, producing arsenates.

Photosystems are arrangements of chlorophyll and other pigments packed into thylakoids. Many Prokaryotes have only one photosystem, Photosystem II (so numbered because, it was the second one discovered, though it was most likely the first to evolve). Eukaryotes have Photosystem II and Photosystem I. Photosystem I uses chlorophyll-a, in the form referred to as P700. Photosystem II uses a form of chlorophyll-a known as P680.

Both "active" forms of chlorophyll-a function in photosynthesis due to their association with proteins in the thylakoid membrane.

### 2.6.1 Light Reactions

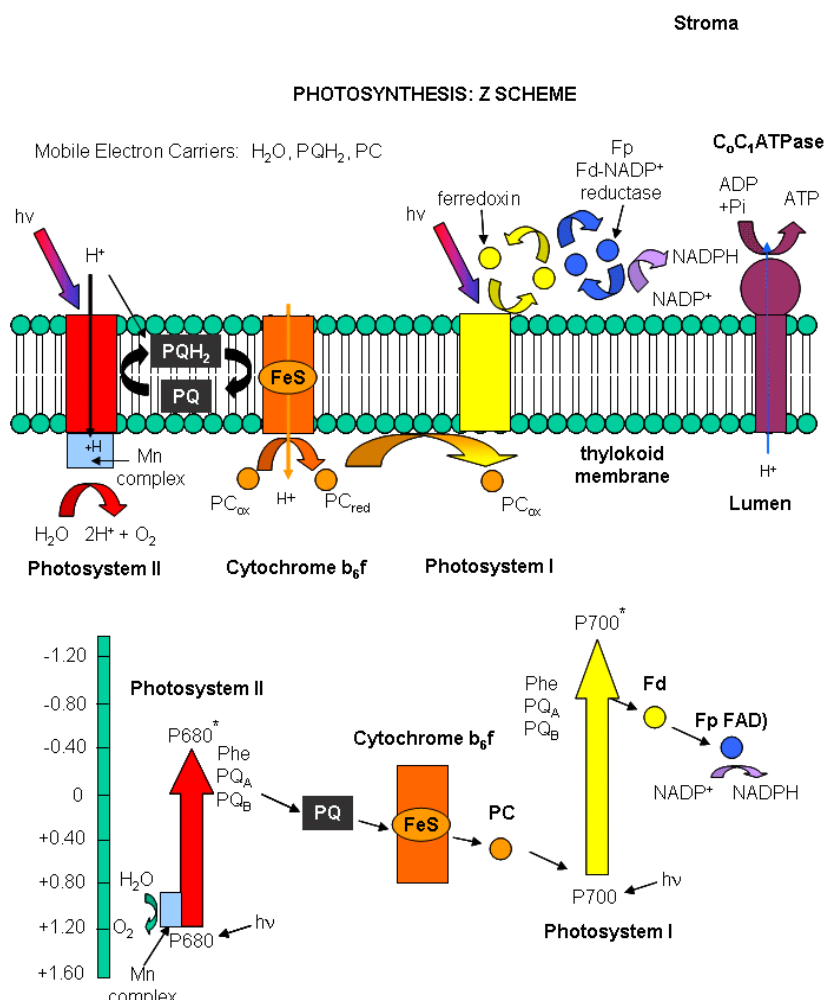
The Light Dependent Process (Light Reactions) requires the direct energy of light to make energy carrier molecules and takes place in and around the thylakoid membranes. In the Light Dependent Processes, light strikes chlorophyll-a in such a way as to excite electrons to a higher energy state. In a series of reactions, the energy is converted (along an electron transport process) into ATP and NADPH. Water is split in the process, releasing oxygen as a by-product of the reaction. The ATP and NADPH are used to make C-C bonds in the Light Independent Process (Dark Reactions). The Z-scheme (Figure 13) represents the steps in the light reactions, showing the pathway of electron transport from water to NADP<sup>+</sup> (nicotinamide adenine dinucleotide phosphate).

The light reaction occurs in two photosystems. Light energy absorbed by photosystem II causes the formation of high-energy electrons, which are transferred along a series of acceptor molecules in an electron transport chain to photosystem I. Photosystem II obtains replacement electrons from water molecules, resulting in their split into hydrogen ions (H<sup>+</sup>) and oxygen atoms. The oxygen atoms combine to form molecular oxygen (O<sub>2</sub>), which is released into the atmosphere. The hydrogen ions are released into the lumen. Additional hydrogen ions are pumped into the lumen by electron acceptor molecules, creating a high concentration of ions inside the lumen.

The flow of hydrogen ions back across the photosynthetic membrane provides the energy needed to drive the synthesis of the energy-rich molecule ATP. High-energy electrons released as photosystem I absorbs light energy are used to drive the synthesis of NADPH. Photosystem I obtains electrons from the electron transport chain. ATP provides the energy and NADPH provides the hydrogen atoms needed to drive the subsequent photosynthetic dark reaction, or the Calvin cycle.



## INTRODUCTION



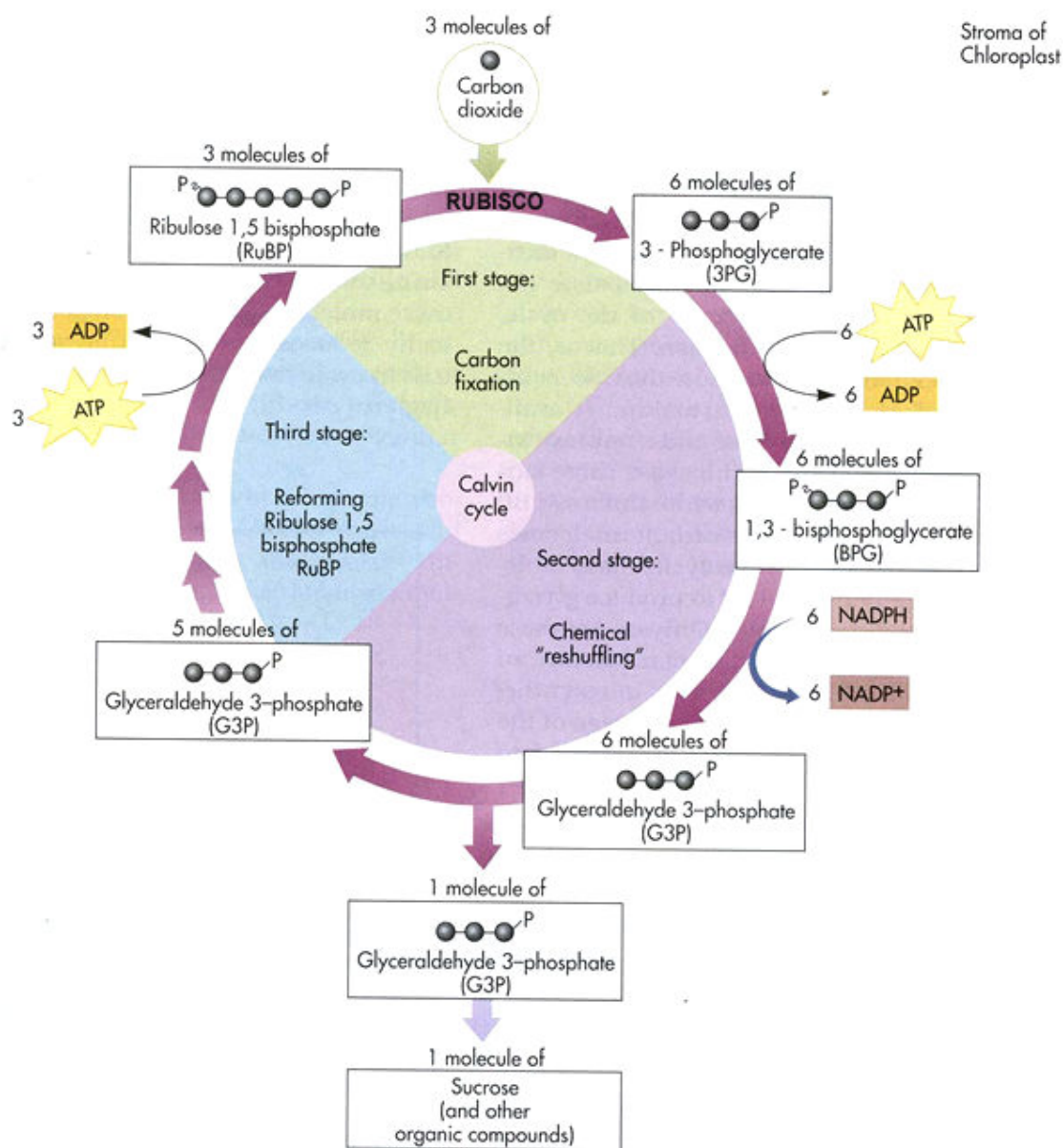
**Figure 13. The Z scheme.**

The Z-scheme represents the steps in the light reactions, showing the pathway of electron transport from water to  $\text{NADP}^+$  (nicotinamide adenine dinucleotide phosphate). This leads to the release of oxygen, the "reduction" of  $\text{NADP}^+$  to  $\text{NADPH}$  (by adding two electrons and one proton), and the building-up of a high concentration of hydrogen ions inside the thylakoid lumen (needed for ATP production).

### 2.6.2 Dark Reaction

In the Light Independent or Dark reaction, carbon dioxide from the atmosphere (or water for aquatic/marine organisms) is captured and modified by the addition of hydrogen, to form carbohydrates. The incorporation of carbon dioxide into organic compounds is known as carbon fixation. These reactions occur in the stroma matrix and are also called the Calvin-Benson-Bassham cycle, photosynthetic carbon reduction cycle, reductive

pentose phosphate pathway or C3 cycle (Figure 14). The energy for this reaction comes from the light reactions.



**Figure 14. The Calvin cycle**

The Calvin cycle can be explained in three phases. In phase 1 (Carbon Fixation), the enzyme Rubisco catalyses the incorporation of  $\text{CO}_2$  into a five-carbon sugar, ribulose bisphosphate (RuBP). The product of the reaction is a six-carbon intermediate which immediately splits in half to form two molecules of 3-phosphoglycerate. In phase 2 (Reduction), ATP and  $\text{NADPH}_2$  from the light reactions are used to convert 3-phosphoglycerate to glyceraldehyde 3-phosphate, the three-carbon carbohydrate precursor to glucose and



other sugars. In phase 3 (Regeneration), more ATP is used to convert some of the pool of glyceraldehyde 3-phosphate back to RuBP, the acceptor for CO<sub>2</sub>, thereby completing the cycle. For every three molecules of CO<sub>2</sub> that enter the cycle, the net output is one molecule of glyceraldehyde 3-phosphate (G3P). For each G3P synthesized, the cycle spends nine molecules of ATP and six molecules of NADPH<sub>2</sub>. The light reactions sustain the Calvin cycle by regenerating the ATP and NADPH<sub>2</sub> (Calvin, 1989).

CO<sub>2</sub> comes into the stroma of the chloroplast via the stomata of the leaves. The Calvin cycle (Figure 14) begins with the enzyme Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase) catalyzing the carboxylation of Ribulose-1,5-bisphosphate, a 5 carbon compound, by carbon dioxide (a total of 6 carbons) in a two-step reaction. The initial product of the reaction is a six-carbon intermediate so unstable that it immediately splits in half, forming two molecules of glycerate 3-phosphate, a 3-carbon compound (also: 3-phosphoglycerate, 3-phosphoglyceric acid, 3PGA). The PGA molecules are further phosphorylated (by ATP produced during light reaction) and are reduced (by NADPH produced during light reaction) to form phosphoglyceraldehyde (PGAL). PGAL serves as the starting material for the synthesis of glucose and fructose. Glucose and fructose make the disaccharide sucrose, which travels in solution to other parts of the plant (e.g., fruit, roots). Glucose is also the monomer used in the synthesis of the polysaccharides starch and cellulose. Majority of the PGAL is recycled and turned back into RuBP so that the cycle can continue (Figure 14). Overall, thirteen enzymes are required to catalyze the reactions in the Calvin cycle. The reactions do not involve energy transduction, but rather the rearrangement of chemical energy. Each molecule of CO<sub>2</sub> reduced to a sugar [CH<sub>2</sub>O]<sub>n</sub> requires 2 molecules of NADPH and 3 molecules of ATP.

### **2.7 Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase)**

The enzyme Rubisco is found in most autotrophic organisms from prokaryotes (photosynthetic and chemoautotrophic bacteria, cyanobacteria and archaea) to eukaryotes (various algae and higher plants) (Andersson and Backlund, 2008). Rubisco catalyzes the initial step in CO<sub>2</sub> fixation, the carboxylation of ribulose 1, 5-bisphosphate (RuBP), and yielding two molecules of phosphoglycerate. RuBP, the initial CO<sub>2</sub> acceptor, is regenerated in the Calvin cycle, and the fixed carbon is incorporated into carbohydrates such as sucrose and starch. By catalyzing the photosynthetic fixation of carbon dioxide,

this enzyme is the agent responsible for bringing into organic combination virtually all the carbon atoms found in living organisms. Despite this vital role, Rubisco is a poor catalyst, having both a low affinity for carbon dioxide and a small turnover number (3 per second); thus autotrophic organisms devote a major part of their synthetic effort to making many molecules of the enzyme. Indeed, Rubisco is regarded as the most abundant single protein in the biosphere (Ellis, 1979). It is certainly the most abundant enzyme: up to 50% of leaf proteins in plants are Rubisco. This large amount does not result primarily from the enormous task the enzyme has to carry out, but rather reflects the catalytic inefficiency of Rubisco as a catalyst.

Not only is the enzyme slow, it also catalyzes a competing oxygenase reaction that leads to loss of energy by photorespiration. In this reaction,  $O_2$  instead of  $CO_2$  is added to RuBP, yielding phosphoglycolate and 3-phosphoglycerate. Phosphoglycolate is metabolized in the glycolate pathway which involves enzymes and cytochromes located in the mitochondria and peroxisomes. In this process, two molecules of phosphoglycolate are converted to one molecule of carbon dioxide and one molecule of 3-phosphoglycerate, which can reenter the Calvin cycle. Some of the phosphoglycolate entering this pathway can be retained by plants to produce other molecules such as glycine. At air levels of carbon dioxide and oxygen, the ratio of the reactions is about 4 to 1, which results in a net carbon dioxide fixation of only 3.5.

Photorespiration is an energy consuming wasteful process causing a constant drain on the pool of sugar substrate, RuBP and results in a decrease of the efficiency of carbon fixation by up to 50% (Andersson and Taylor, 2003). This photorespiratory pathway severely affects a plant's water-use efficiency and nitrogen budget. It also produces waste ammonia that must be detoxified at a substantial cost to the cell in ATP and reducing equivalents.

Carboxylation and oxygenation of RuBP occur at the same catalytic site of Rubisco; both gaseous substrates compete for the second substrate, RuBP. Therefore, the ratio of carboxylation towards oxygenation is influenced by the relative concentrations of  $CO_2$  and  $O_2$ . Higher  $CO_2$  concentrations result in more efficient photosynthesis with

faster production of biomass. The efficiency with which CO<sub>2</sub> is able to compete with O<sub>2</sub> is quantified by the CO<sub>2</sub>/O<sub>2</sub> specificity factor (referred to as  $\Omega$ ) and is defined as  $V_c K_o / V_o K_c$ , where  $V_c$  and  $V_o$  are the maximal velocities of carboxylation and oxygenation, respectively, and  $K_c$  and  $K_o$  are the Michaelis constants for CO<sub>2</sub> and O<sub>2</sub>, respectively. Thus the relative rates for carboxylation and oxygenation are defined by the product of the specificity factor and the ratio of CO<sub>2</sub> to O<sub>2</sub> concentrations at the active site. The specificity values of Rubisco enzymes from different species and evolutionary lineages differ substantially. Some photosynthesizing bacteria of the  $\alpha$ -proteobacteria group have the lowest specificity values (5-40) whereas members of the Rhodophyta (red algae) have the highest (180-240). Chlorophyta, such as higher plants and green algae have intermediate specificity values in the range of 60-100. An inverse correlation between specificity and turnover rate ( $V_c$  or  $k_{cat}$  for carboxylation) has been observed with e.g. bacteria displaying low specificity values and high turnover rates whereas higher plants have high specificity values coupled to low turnover rates. In addition, intracellular CO<sub>2</sub> and O<sub>2</sub> concentrations vary considerably among species because several organisms, including plants, have evolved mechanisms (carboxysomes, pyrenoids, C<sub>4</sub>- and CAM metabolisms) that concentrate CO<sub>2</sub> at the carboxylation site (Andersson and Backlund, 2008).

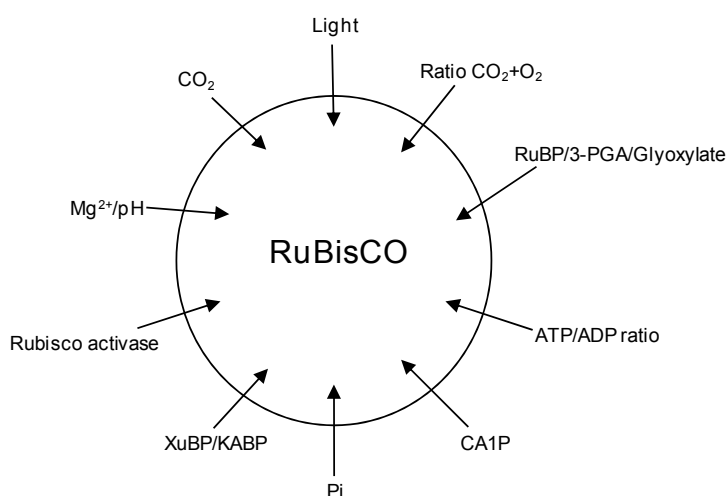
Rubisco is slow, being able to fix only 3 carbon dioxide molecules each second. Nevertheless, because of its extremely large concentration, under most conditions, and when light is not otherwise limiting photosynthesis, the reaction of Rubisco responds positively to increasing carbon dioxide concentration, therefore the concentration of carbon dioxide is limiting. The ultimate rate-limiting factor of the Calvin cycle is Rubisco that cannot be ameliorated in short time by any other factor. For these reasons, genetic redesign of Rubisco with the aim of constructing transgenic plants with improved photosynthetic efficiency and thereby increased agricultural productivity has attracted a lot of interest (Schneider et al., 1992).

### 2.7.1 Regulation of Rubisco

Rubisco is usually only active during the day, because RuBP is not being produced in the dark, due to the regulation of several other enzymes in the Calvin cycle. In addition, the activity of Rubisco is coordinated with that of the other enzymes of the Calvin cycle in several ways (Figure 15). Upon illumination of the chloroplasts, the pH of the stroma rises from 7.0 to 8.0 because of the proton (hydrogen ion,  $H^+$ ) gradient created across the thylakoid membrane. At the same time, magnesium ions ( $Mg^{2+}$ ) move out of the thylakoids, increasing the concentration of magnesium in the stroma of the chloroplasts. Rubisco has a high optimal pH (can be  $>9.0$ , depending on the magnesium ion concentration) and thus becomes "activated" by the addition of carbon dioxide and magnesium to the active sites (Figure 15). Rubisco and other enzymes involved in the cycle are affected by the oxidation-reduction state of thioredoxin. In chloroplasts, ferredoxin reduces thioredoxin. Glyceraldehyde 3-phosphate dehydrogenase and ribulose 5'-phosphate kinase are regulated directly by NADPH, as well as by thioredoxin.

A common feature of all Rubisco molecules is a chemical modification step necessary to convert the enzyme from its inactive to its active form. The activation process consists of the formation of a carbamate group by reaction of a  $CO_2$  molecule with the  $\epsilon$ - amino group of a lysine residue (Lys 201) at the active site. This activator  $CO_2$  molecule is separate from the  $CO_2$  molecule that becomes incorporated into RuBP during catalysis. Formation of the carbamate is followed by rapid binding of  $Mg^{2+}$ , resulting in the active ternary complex-enzyme,  $CO_2$ , and  $Mg^{2+}$  (Schneider et al., 1992). In plants and some algae, another enzyme, Rubisco activase is required to allow the rapid formation of the critical carbamate in the active site of Rubisco {Weiner, 1994 #454}. Activase is required because RuBP substrate binds more strongly to the active sites lacking the carbamate and markedly slows down the "activation" process. In the light, Rubisco activase promotes the release of the inhibitory or storage RuBP from the catalytic sites. In some plants (e.g. tobacco and many beans), in the dark, Rubisco is inhibited by a competitive inhibitor/ a substrate analog 2-Carboxy-D-arabitol 1-phosphate (CA1P) synthesized by these plants (Figure 15). CA1P binds tightly to the active site of

carbamylated Rubisco and inhibits catalytic activity. During dark and in the light, Rubisco activase promotes the release of CA1P from the catalytic sites (Lilley and Portis, 1990). After the CA1P is released from Rubisco, it is rapidly converted to a non-inhibitory form by a light-activated CA1P-phosphatase. Once every several hundred reactions, the normal reactions with carbon dioxide or oxygen are not completed and other inhibitory substrate analogs are formed in the active site. Rubisco activase can promote the release of these analogs from the catalytic sites and maintain the enzyme in a catalytically active form. In the initial reaction of Rubisco in the light, the RuBP that was separated from Rubisco binds with the carbamylated enzyme and after proton abstraction produces Enediol that can react with carbon dioxide. A limitation of either Rubisco or RuBP at any stage will make the reaction insensitive to any other factor including carbon dioxide.



**Figure 15. Multiple factors involved in regulating the activity of Rubisco.**

Many factors like light, CO<sub>2</sub>/O<sub>2</sub> ratio, Rubisco activase, several enzymes of the Calvin cycle regulate the activity of Rubisco.

### 2.7.2 Rubisco Structure

Rubisco in many organisms is composed of two types of subunits, the catalytic large (RbcL, 50 to 55 kDa), and the small (RbcS, 12 to 18 kDa) subunits. Based on the presence or absence of the small subunit and also on the primary sequence of the large subunit, four forms or types of Rubisco have been distinguished (Li et al., 2005; Tabita, 1999). Of the four forms of Rubisco, form I is the most abundant, occurring in most

chemoautotrophic bacteria, cyanobacteria, red and brown algae, and in all higher plants. It is composed of eight large and eight small subunits arranged in a hexadecameric structure,  $\text{RbcL}_8\text{S}_8$  (Baker et al., 1977); with a core of four  $\text{RbcL}_2$  dimers arranged around a four-fold axis, capped at each end by four small subunits (Figure 16 B and C). Based on amino acid sequences of the form I enzymes, a distinction has been made between green-type enzymes (forms I A and B from cyanobacteria, eukaryotic algae and higher plants) and red-type enzymes (forms I C and D from non-green algae and phototropic bacteria) (Andersson, 2008).

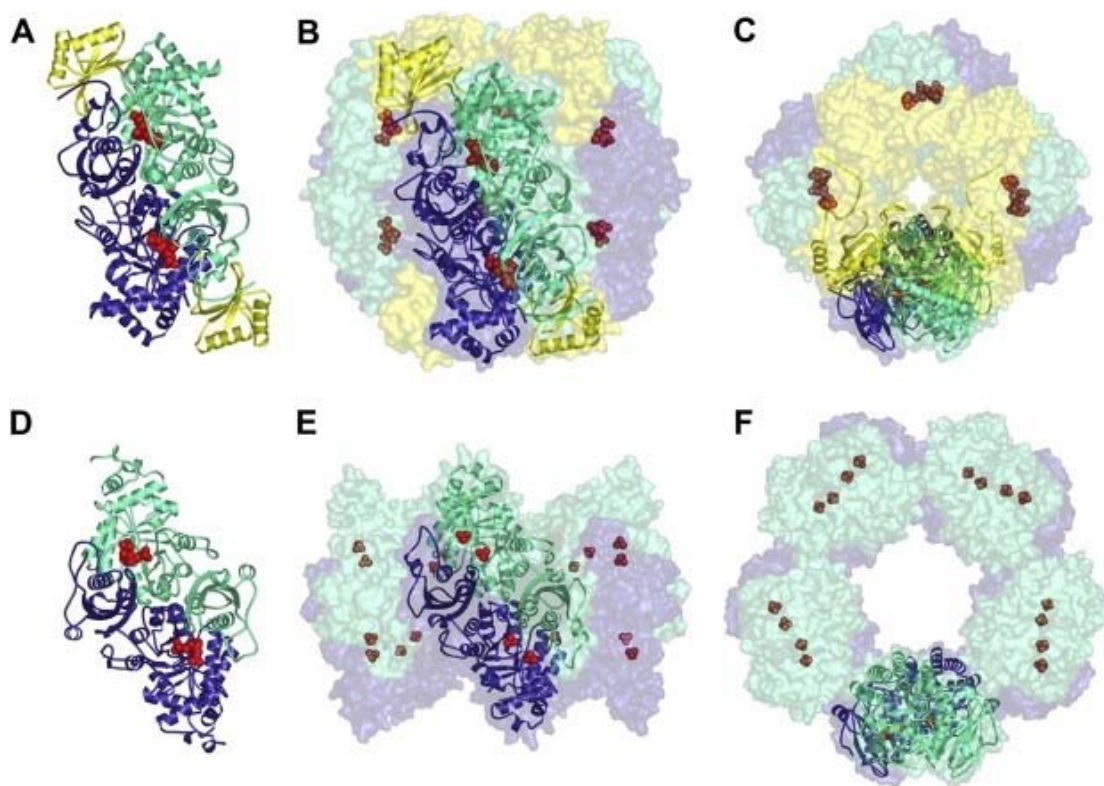
The form II enzyme is a dimer of large subunits,  $\text{Rbc}(\text{L}_2)_n$ , and lacks small subunits. The form II enzyme was initially discovered in purple non-sulphur bacteria, *Rhodospirillum rubrum*, but has also been found in several chemoautotrophic bacteria and in eukaryotic dinoflagellates. The first crystal structure of Rubisco was from *Rhodospirillum rubrum* (Figure 16D) and reveals high similarity to the large subunit structure of form I Rubisco (Andersson and Backlund, 2008). Several non-sulphur phototropic bacteria, i.e. *Rhodobacter sphaeroides*, *R. capsulatus*, several *Thiobacillus* sp., and *Hydrogenovibrio marinus* contain both form I and form II enzymes.

Form III Rubisco is found only in archaea, and has been shown to form either dimers ( $\text{RbcL}_2$ ) (Finn and Tabita, 2003; Watson et al., 1999) or decamers ( $[\text{RbcL}_2]_5$ ) (Maeda et al., 1999), depending on the organism. The crystal structure of the form III Rubisco from *Thermococcus kodakaraensis* (Figure 16 E and F) reveals that the protein is comprised of a pentamer of dimers (Kitano et al., 2001). Its dimeric interface is very similar to those observed in the large subunit of form I and form II Rubisco.

Form IV, also called the Rubisco-like protein (RLP), was recently discovered to be a homolog of Rubisco (Hanson and Tabita, 2001; Li et al., 2005). Members of the form IV subfamily do not catalyze the carboxylation reaction, but have been shown to play a role in sulphur metabolism. RLPs from species of Bacilli (*Bacillus subtilis* and *Geobacillus kaustophilus*) have been found to catalyze enolisation of 2,3-diketo-5-methyl-thiopentyl-1-phosphate, a compound with structural similarity to RuBP. A unique feature of *M. aeruginosa* is that it also harbours a gene encoding the form I Rubisco large subunit in



addition to form IV. Crystal structures have been reported for the RLPs from *Chlorobium tepidum* (green sulphur bacterium), *Geobacillus kaustophilus*, and *Rhodospseudomonas palustris*. From the phylogenetic analysis, it appears that the RLPs form an entity, and are not dispersed among the true Rubiscos, thus indicating a common origin for either Rubisco or RLPs (Andersson and Backlund, 2008).



**Figure 16. The different arrangement of the quaternary structure of Rubisco enzymes.**

A) The  $RbcL_2S_2$  unit of form I Rubisco from spinach viewed along the 2-fold symmetry axis. Large subunits are blue and green, small subunits are yellow, and the substrate mimic (2CABP) is displayed as red spheres.

B) The entire  $RbcL_8S_8$  hexadecamer along the 2-fold axis and C) along the 4-fold axis.

D) The dimeric form II Rubisco from *Rhodospirillum rubrum* showing the 2-fold symmetry.

E) and F) The  $RbcL_{10}$  Rubisco from *Thermococcus kodakaraensis* viewed along the 2-fold and 5-fold axes, respectively. Sulphate ions bound in the active site are displayed as red spheres (Andersson and Backlund, 2008).

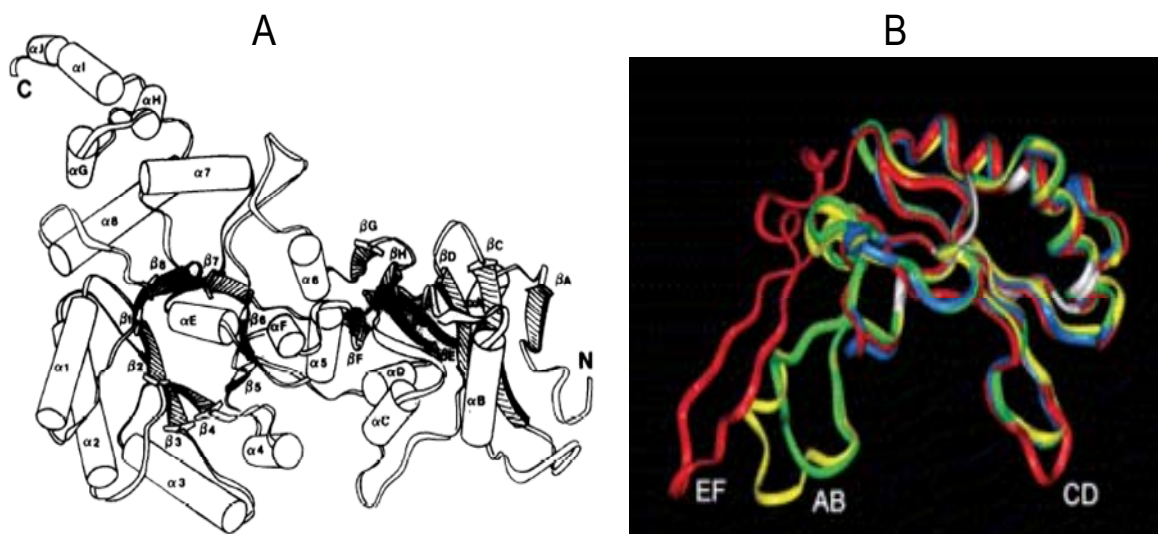
Despite rather low sequence identity (25–30%), the hexadecameric (Form I) RbcL subunits and the dimeric (Form II) RbcL subunits have very similar three dimensional structures. The overall fold of the large subunit is similar in all forms of Rubisco consisting of a smaller amino-terminal domain consisting of a four-to-five-stranded mixed  $\beta$  sheet with helices on one side of the sheet and a larger carboxy-terminal domain (Figure 17A). The carboxy-terminal domain consists of eight consecutive  $\beta\alpha$ -units arranged as an eight-stranded parallel  $\alpha/\beta$  barrel structure. The active site is located at the carboxy-terminal end of the  $\beta$ -strands, with the loops connecting the  $\beta\alpha$ -units contributing several residues involved in catalysis and substrate binding. Residues from the amino-terminal domain of the adjacent large subunit in the dimer complete the active site. Thus, the functional unit structure of Rubisco is an RbcL<sub>2</sub> dimer of large subunits harbouring two active sites (Andersson and Backlund, 2008).

In the RbcL<sub>8</sub>S<sub>8</sub> enzymes, four RbcL<sub>2</sub> units are arranged around a fourfold symmetry axis with the entrances to the eight active sites facing the outside of the hexadecameric molecule (Figure 16 B and C). A RbcS subunit is associated with each RbcL subunit, also obeying these twofold and fourfold symmetries and interacting with three large subunits. A solvent channel traverses the molecule along its fourfold axis. Binding of the substrate or inhibitors to the non-activated enzyme locks the enzyme in a closed unproductive form. In the inactive enzyme, the active site is open and accessible to activating cofactors and bisphosphate substrate (Curmi et al., 1992). After formation of the essential carbamate and coordination of the Mg<sup>2+</sup>, RuBP substrate binds and a series of loops close over the site to enfold and capture the bisphosphate (Knight et al., 1990; Newman and Gutteridge, 1993). Closure of the loops brings together amino acids that are critical for catalysis and determine the fate of the substrate.

The RbcS subunits are not absolutely required for activity, as indicated by the existence of the dimeric Form II and the decameric archeal Rubisco enzymes. However the presence of RbcS subunits improves catalysis in some rather specific, but undefined way. The RbcS subunits are more divergent than the L subunits, both in sequences and in three-dimensional structures. The common core structure consists of a four-stranded



anti-parallel  $\beta$ -sheet covered on one side by two helices (Figure 17B). The most striking variations occur in two distinct locations, the loop between  $\beta$  strands A and B of the small subunit, the so-called  $\beta$ A- $\beta$ B-loop and the carboxy-terminus. The  $\beta$ A-  $\beta$ B-loops of four small subunits line the openings of the solvent channel that traverses the holoenzyme. Rubisco from prokaryotes and non-green algae have only ten residues in the loop as illustrated by the structure of the cyanobacterial enzyme, but Rubisco from higher plants have 22 and green algal Rubisco have 28. Non-green algae and some prokaryotes, which have ten residues in their  $\beta$ A-  $\beta$ B-loops, display carboxy-terminal extensions that form  $\beta$  hairpin structures ( $\beta$ E- $\beta$ F loop) in the spaces that are normally occupied by the longer  $\beta$ A- $\beta$ B-loops of the green algal and plant enzymes. Four  $\beta$ -hairpin structures form a central  $\beta$ -barrel at the entrance to the central solvent channel. These enzymes also display a slightly longer loop between  $\beta$ strands C and D (Figure 17B). The small subunits of green algae also have longer carboxy-termini than those of higher plants, but these do not form  $\beta$ -hairpins and do not appear to be essential to the function of the enzyme. The small subunits covering a substantial area at two opposite ends of the RbcL-subunit octamer may be necessary to assemble and concentrate the large catalytic subunits. Since Rubisco enzymes lacking small subunits have the lowest specificity values, they might contribute to the differences in kinetic properties observed among different Rubisco enzymes (Andersson and Backlund, 2008).



## Figure 17. Rubisco large and small subunits.

A. Schematic view of the large subunit from *Rhodospirillum rubrum* with  $\alpha$  subunits represented by cylinders and  $\beta$ -strands by arrows (Schneider et al., 1990).

B. Comparison of the Rubisco small-subunit X-ray crystal structures from *Chlamydomonas reinhardtii* (yellow), spinach shown in green, *Synechococcus* sp. PCC6301 (blue), and *Galdieria partita* (red). Loops are labeled relative to their flanking  $\beta$  strands. Residues that are more than 95% conserved among all known small-subunit sequences are colored white (Spreitzer, 2003).

### 2.7.3 Synthesis, Folding and assembly of Rubisco

Synthesis and assembly of functional Rubisco in plants and green algae require communication between organelles, because RbcS subunits are encoded by the nuclear genome and synthesized in the cytosol, whereas RbcL subunits are encoded by the chloroplast genome and synthesized on chloroplast ribosomes. The *rbcL* gene in higher plants is present as a single copy per chloroplast genome, but because many copies of the genome are present in each plastid, the actual *rbcL* copy number per chloroplast can be high. With rare exceptions, *rbcL* in higher plants does not contain introns and encodes ~ 475 amino acids (Gutteridge and Gatenby, 1995). Control of the expression of genes for Rubisco occurs both transcriptionally and posttranscriptionally, but apparently differs in the nucleus as compared to the chloroplast. After translation, newly synthesized S subunits must be translocated across the chloroplast membrane, where an N-terminal signal peptide is proteolyzed prior to assembly with RbcL subunits. Proteolysis of a small N-terminal region of nascent L subunit is also observed. Covalent posttranslational modifications, including acetylation, N-methylation, phosphorylation, and possibly transglutamination, have been noted in plant and cyanobacterial Rubiscos. Assembly of both RbcL<sub>2</sub> and RbcL<sub>8</sub>S<sub>8</sub> Rubisco from photosynthetic prokaryotes and rhodophytic algae is apparently simpler, as the genes encoding their subunits are found within the same operon and the synthesis and assembly occurs in the cytosol (Hartman and Harpel, 1994).

Chloroplast homologs of DnaK, DnaJ, and GrpE have been reported (Checa and Viale, 1997). Whereas the RbcLS of Rubisco has not been found in direct association with these chaperones, other polypeptides that must be imported into the chloroplast,

including the transit peptide of the small subunit have been shown to be associated. Moreover, the small subunit transit peptide has been shown to interact directly with chloroplast Hsp70 during import and to contain a recognition domain for Hsp70. The requirement for DnaK and DnaJ could be partially overcome by overexpressing the GroEL and GroES proteins.

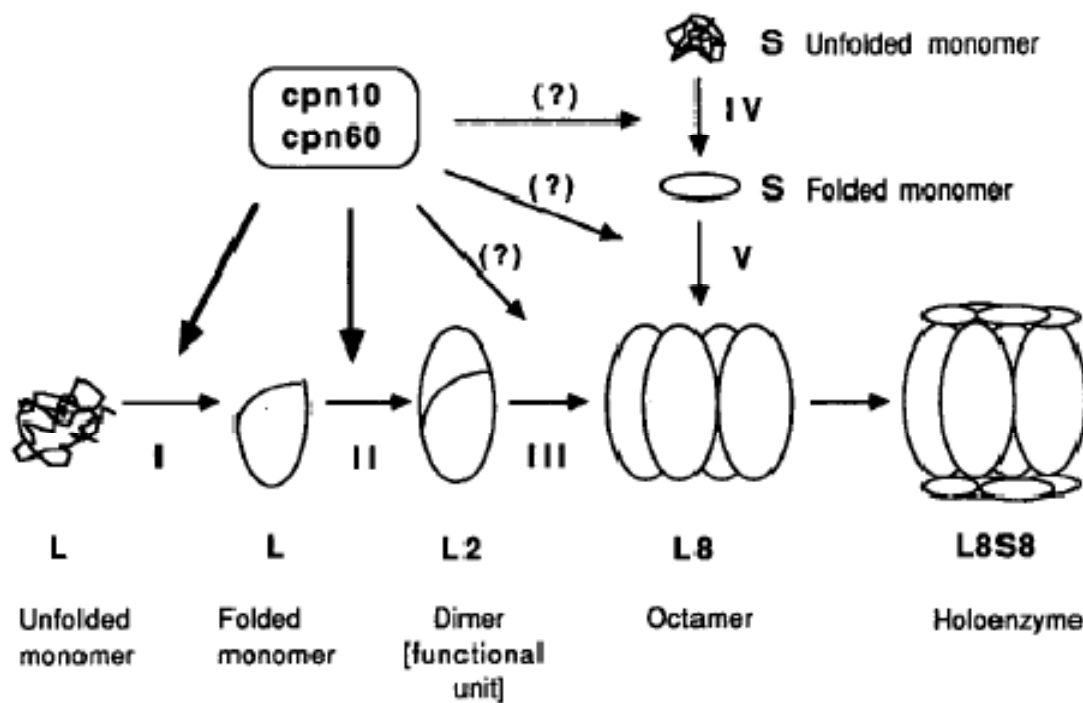
The complexes formed between the newly translated polypeptides and the DnaK/DnaJ/GrpE complexes represent the first step in cooperative protein folding that ends with the release from the DnaK/DnaJ/GrpE complexes to the GroEL/ES chaperonins. A similar set of reactions probably operates in chloroplasts but with regard to the RbcLS of Rubisco, more information exists at the level of GroEL/ES folding than at the earlier steps in protein folding. Evidence for chaperonin-mediated protein folding in the chloroplast was found over 2 decades ago when studies revealed that newly synthesized Rubisco peptides were transiently associated with a large multimeric complex prior to appearance in the holoenzyme through an ATP-dependent reaction (Ellis, 1991; Milos, 1984).

It is now well documented that the assembly of Rubisco requires chloroplast homologs of the GroEL and GroES chaperonins (Gatenby and Ellis, 1990; Goloubinoff et al., 1989a; Gutteridge and Gatenby, 1995) and that many if not all chloroplast-imported proteins interact with these proteins.

With the help of the GroEL/ES chaperone system, *in vivo* and *in vitro* refolding of dimeric form II Rubisco from *R. rubrum* have been successful (Goloubinoff et al., 1989a). However, very limited success has been achieved for the reconstitution of Form I Rubisco. While GroEL/ES dependent *in vivo* expression of active Rubisco holoenzyme from *Synechococcus* sp. PCC6301 was successful (Goloubinoff et al., 1989b), the *in vivo* attempts for the assembly of higher-plant Rubisco failed (Andrews and Lorimer, 1985).

From the experiments on cyanobacterial Rubisco, the current Rubisco assembly model (Figure 18) states that assembly proceeds through dimeric (RbcL<sub>2</sub>) and octameric (RbcL<sub>8</sub>) intermediates, and that small subunits bind in the final stages (Gatenby et al.,

1988; Goloubinoff et al., 1989a; Lee and Tabita, 1990); i.e. the folded large subunits dimerize into  $\text{RbcL}_2$ , the dimers form tetramers leading to a core of 8 large subunits ( $\text{RbcL}_8$  core) to which the folded small subunits associate, resulting in the  $\text{RbcL}_8\text{S}_8$  type of active holoenzyme.



**Figure 18. Proposed assembly pathway for Rubisco hexadecamers.**

The  $\text{L}_2$  dimer is the basic structural motif. The assembly of the hexadecameric  $\text{L}_8\text{S}_8$  type of holoenzyme starts from the folding of large subunit monomer, formation of the  $\text{L}_2$  dimer, tetramerization of 2 dimers to give the  $\text{L}_8$  core and then the association of folded small subunit monomers with the  $\text{L}_8$  core (Goloubinoff et al., 1989b).

## 2.8 RbcX

In some cyanobacteria, a gene called *rbcX* is present between *rbcL* and *rbcS* and is cotranscribed with the *rbcL* and *rbcS* genes (Larimer and Soper, 1993). The RbcX product (~15.5 kDa) of the intermediary *rbcX* gene is not part of the final Rubisco complex

and, unlike RbcL and RbcS whose sequences are highly conserved by functional constraints, the RbcX sequence is highly variable (56% similarity) amongst cyanobacterial species (Rudi et al., 1998). Nevertheless, the juxtaposition of *rbcX* within an *rbcLXS* operon is highly conserved in  $\beta$ -cyanobacteria, suggesting that the RbcX product may function in a role associated with CO<sub>2</sub> fixation (Larimer and Soper, 1993; Li and Tabita, 1997; Onizuka et al., 2004). RbcX from the cyanobacteria *Anabaena* sp. CA and *Synechococcus* sp. PCC7002 has been found to increase the amount of functional Rubisco correctly assembled in *E. coli*, consistent with a molecular chaperone-like function (Li and Tabita, 1997; Onizuka et al., 2004).

In some cyanobacteria like *Synechococcus* sp. PCC7942, *rbcX* is present outside the operon for *rbcLS* and RbcX has been shown to be not affecting the assembly and hence the activity of its Rubisco recombinantly expressed in *E. coli* (Emlyn-Jones et al., 2006).

### 3 Aim of the study

Rubisco is the most abundant protein on earth. Rubisco catalyses the initial step of carbon dioxide fixation in photosynthesis and so is crucial for the agronomic performance of plants. Despite this vital role, Rubisco is inefficient as a catalyst because it participates not only in photosynthesis, but also in photorespiration, which is considered as a wasteful process (Tabita 1999). Efficiency of Rubisco varies among different photosynthetic organisms. Rubisco from red algae possess higher CO<sub>2</sub> specificity compared to higher plants (Andersson and Backlund, 2008). *in vitro* mutagenesis of Rubisco has not been successful since all the attempts to produce active form I Rubisco from many cyanobacteria and higher plants, *in vivo* or *in vitro*, failed so far (Chaudhari and Roy, 1989; Cloney et al., 1993; van der Vies et al., 1986).

Rubisco is one of the stringent substrates of GroEL (Brinker et al., 2001). Similar to the bacterial GroEL/ES system, the chloroplast homologue ch-cpn60/20 system is implicated in the efficient folding of Rubisco (Dickson et al., 2000; Gatenby and Ellis, 1990; Goloubinoff et al., 1989a; Goloubinoff et al., 1989b; van der Vies et al., 1986). While the structural details for GroEL/ES are well established (Langer et al., 1992b; Saibil et al., 1991; Chandrasekar, 1986; Tilly et al., 1981), a detailed structural analysis of ch-cpn60 is yet to be achieved. In contrast to GroEL, the ch-cpn60 has two types of cochaperone, cpn10 and cpn20, the latter consisting of a tandem repeat of cpn10 units (Hill and Hemmingsen, 2001). Unlike the bacterial GroES and the chloroplast cpn10, which are heptameric (Sharkia et al., 2003), the chloroplast cpn20 has been reported to form tetramers (Koumoto et al., 1999). One of the aims of this study was to structurally characterize chloroplast cpn60/20 complexes, using electron microscopy, to understand the nature of interaction of ch-cpn20 with the ch-cpn60.

Recently, it has been reported that the *rbcX* gene of cyanobacteria enhances the production of enzymatically active Rubisco upon coexpression with *rbcL* and *rbcS* in *E. coli* (Emlyn-Jones et al., 2006; Li and Tabita, 1997; Onizuka et al., 2004). The structural

analysis of RbcX from *Synechococcus* sp. PCC7002, using X-ray crystallography, demonstrated that the RbcX dimer functions as an assembly chaperone for the hexadecameric Rubisco (Saschenbrecker et al., 2007). These studies revealed that the mutations disrupting the conserved peripheral polar surface, (Q29A) or (R70A), or those disrupting the central crevice (Y17A,Y20L) in the RbcX, cause a defect in the RbcX function (Saschenbrecker et al., 2007). The structure of RbcX mutants was to be analyzed by X-ray crystallography to show that the functional defects of RbcX mutants were not a consequence of protein misfolding.

RbcX homologues and the RbcX interacting C-terminal peptide sequence of RbcL (Saschenbrecker et al., 2007) exist in higher plants as well. Consequently, another objective of this study was to characterize RbcX of *A. thaliana*.

## 4 Materials and methods

### 4.1 Laboratory equipment

**Abimed** (Langenfeld, Germany): Gilson Pipetman (2 to 1000 µl).

**Amersham Pharmacia Biotech** (Freiburg, Germany): ÄKTA Explorer; SMART-System; chromatography columns: Mono-Q, HiTrap-Heparin, Superdex 200, Superose 6; resins: Source 30-Q; electrophoresis power supplies: EPS200, EPS600.

**Amicon** (Beverly, MA, USA): concentration chambers: Centriprep, Centricon.

**Avestin** (Mannheim, Germany): EmulsiFlex C5 homogenizer.

**Beckmann Coulter GmbH** (Krefeld, Germany): centrifuges (J6-MI, GS-6R, Avanti 30, Avanti J-25I, Optima LE-80K ultracentrifuge), spectrophotometers (DU640, DU800), LS 6500 multi-purpose scintillation counter.

**Biometra** (Göttingen, Germany): PCR-Thermocycler.

**Bio-Rad** (Munich, Germany): electrophoresis chambers: MiniProtean 2 and 3, electrophoresis power supply Power PAC 300.

**Branson** (Connecticut, USA): Sonifier cell disruptor B15.

**Eppendorf** (Hamburg, Germany): centrifuges: 5415C, 5417R; Thermomixer Comfort.

**Forma Scientific** (Marietta / OH, USA): Orbital Shaker 4581

**Fuji** (Tokyo, Japan): Phosphoimager FLA-2000; ImageReader LAS-3000.

**Gefinge** (Gefinge, Sweden): autoclave.

**Hoefer Scientific Instruments** (San Francisco, USA): SemiPhore blotting transfer unit.

**HORIBA Jobin Yvon GmbH** (München, Germany): Spex Fluorolog 3.

**Mettler Toledo** (Giessen, Germany): balances: AE160, AG285, PB602.

**Millipore** (Eschborn, Germany): deionization system MilliQ plus PF; Millex-HA filters (0.22 µm); vacuum filtration unit (0.22 µm).

**MWG BiotechAG** (Göttingen, Germany): gel documentation system BioCapt.

**New Brunswick Scientific** (Nürtingen, Germany): orbital shaker and incubator Innova 4430.

**Philips** (Amsterdam, Netherlands): electron microscope CM12.

**Raytest** (Straubenhardt, Germany): AIDA gel imaging software version 3.5.



**Savant/Thermoquest** (Engelsback, Germany): Stacked Gel Dryer SGD300.

**WTW** (Weilheim, Germany): pH-Meter.

**Wyatt Technology** (Santa Barbara/CA, USA): FFF-MALS system, software ASTRA.

## 4.2 Materials

### 4.2.1 Chemicals

Chemicals and biochemicals used in this work were of *pro analysi* grade and purchased from Fluka (Deisenhofen, Germany), Calbiochem (Bad Soden, Germany), Merck (Darmstadt, Germany), Sigma-Aldrich (Steinheim, Germany), Roth (Karlsruhe, Germany), and Roche (Mannheim, Germany) unless stated otherwise.

**Amersham Pharmacia Biotech** (Freiburg, Germany): western blotting detection systems: ECL, chromatographic resins, [<sup>35</sup>S]-Methionine, NaH<sup>14</sup>CO<sub>3</sub>.

**BioMol** (Hamburg, Germany): IPTG, HEPES.

**BioRad** (Munich, Germany): ethidiumbromide; Bradford Protein-Assay.

**Cambrex Bio Science** (Rockland / ME, USA): Sea Kem LE Agarose.

**Difco** (Heidelberg, Germany): Bacto tryptone, Bacto yeast extract, Bacto agar.

**Fermentas** (St. Leon-Rot, Germany): GeneRuler 1kb DNA Ladder.

**Fluka** (Deisenhofen, Germany): acetic acid, acetone, Bis-Tris, H<sub>2</sub>O<sub>2</sub>, luminol, PEG.

**Hampton Research** (Aliso Viejo / CA, USA): Crystallization screens.

**Invitrogen** (Karlsruhe, Germany): protein marker for SDS PAGE, dNTP set.

**J.M. Gabler Saliter GmbH & Co. KG** (Obergünzburg, Germany): skim milk powder.

**New England Biolabs** (Frankfurt a. Main, Germany): restriction enzymes; calf intestinal alkaline phosphatase (CIP); T4 DNA ligase; prestained protein marker for SDS PAGE.

**Promega** (Mannheim, Germany): Wizard SV Gel and PCR Clean-Up System, Wizard Plus SV Miniprep DNA Purification System, PureYield Plasmid Midiprep System.

**Qiagen** (Hilden, Germany): QIAprep Plasmid Mini and Midi kits; QIAquick PCR purification and gel extraction kit; Ni-NTA agarose.

**Roche** (Mannheim, Germany): RTS *in vitro* translation systems: RTS 100 *E. coli* HY Kit, Complete protease inhibitor cocktail, DTT.

**Roth** (Karlsruhe, Germany): ampicillin, glycine, scintillation fluid.

**Schleicher & Schuell** (Dassel, Germany): protran nitrocellulose transfer membrane; fluted paper filter 595-1/2 (270 mm).

**Serva** (Heidelberg, Germany): Acrylamide-Bis, BSA, Coomassie blue G/R, PMSF, SDS.

**USB** (Cleveland, USA): ammonium sulfate, chloramphenicol, EDTA, MES, tricine, urea.

### 4.2.2 Strains

**Novagen** (Darmstadt, Germany): *E. coli* DH5 $\alpha$

**Stratagene** (Heidelberg, Germany): *E. coli* BL21(DE3)

### 4.2.3 Plasmids, DNA and oligonucleotides

**Novagen** (Darmstadt, Germany): pET11a, pET15b, pET28b, pET30b. Plasmid pHUE was a kind gift from Dr. Spencer Whitney, ANU, Canberra, Australia.

**Metabion** (Martinsried, Germany): oligonucleotides (primers).

Plasmids generated during this study will be described below.

### 4.2.4 Enzymes, proteins, peptides and antibodies

**AgriSera** (Vännäs, Sweden): anti-RbcL antibody.

**Amersham Biosciences** (Freiburg, Germany): porcine RNAGuard ribonuclease inhibitor, Protein A sepharose beads.

**Merck** (Darmstadt, Germany): Benzonase.

**New England Biolabs** (NEB, Frankfurt Am Main, Germany): restriction endonucleases, T4 DNA ligase.

**Fermentas** (St. Leon-Rot, Germany): restriction endonucleases.

**JPT Peptide Technologies GmbH** (Berlin, Germany): peptide array.

**MPI for Biochemistry**, Department Cellular Biochemistry (Martinsried, Germany): purified protein stocks of GroEL, GroEL-SR1 (R452E, E461A, S463A, V464A), GroES, DnaK, DnaJ, GrpE, Gp31, yeast mt-Hsp10, Rr-RbcL, Rr-RbcL(K168E), TEV-protease.

**MPI for Biochemistry**, peptide synthesis service (Martinsried, Germany): oligopeptides.

**MPI for Biochemistry** (Martinsried, Germany): antisera against purified Syn7002-RbcX and *E. coli* GroEL (produced in rabbits).

**Pineda-Antikörper-Service** (Berlin, Germany): anti-Cpn60 $\alpha$  antibody, anti Cpn60 $\beta$ -antibody (peptide antibodies, produced in rabbits).

**Promega** (Mannheim, Germany): Pfu DNA polymerase.

**Roche** (Basel, Switzerland): ProteinaseK, shrimp alkaline phosphatase, hexokinase, MDH.

**Sigma-Aldrich** (Steinheim, Germany): BSA, lysozyme, 3x FLAG peptide, mouse monoclonal anti-FLAG M2 antibody, EZview Red ANTI-FLAG M2 Affinity Gel, HRP-conjugated secondary antibodies.

### 4.2.5 Media

**LB medium:** 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, (+15 g/l agar for solid medium).

### 4.2.6 Buffers and stock solutions

**PBS:** 137 mM NaCl, 2.7 mM KCl, 8.4 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4.

**PonceauS:** 0.2% PonceauS, 3% trichloroacetic acid.

**TAE:** 40 mM Tris-Acetate, 1 mM EDTA, pH 8.0.

**TBS:** 50 mM Tris, 150 mM NaCl, pH 7.5.

**TBST:** 0.1% Tween-20 in TBS.

All other buffers and solutions were prepared as convenient stock solutions and either autoclaved or filter sterilized before usage, if applicable.

## 4.3 Molecular biology methods

All experimental methods used were performed according to “Molecular Cloning” (Sambrook, 1989) unless stated otherwise.

### 4.3.1 Plasmid purification

A single *E. coli* colony containing the plasmid of interest was inoculated in LB medium (supplemented with the appropriate antibiotic) and shaken overnight at 230 RPM

at 37 °C. Plasmid DNA was purified using the QIAprep Plasmid kits or Wizard Plus SV Miniprep DNA Purification System, according to the manufacturer`s protocol.

### **4.3.2 DNA analytical methods**

DNA concentrations were measured by UV absorption spectroscopy at  $\lambda=260$  nm. An optical density of  $OD_{260}=1$  corresponds to approximately 50  $\mu\text{g/ml}$  double stranded DNA. The absorbance ratio 260/280 nm for pure DNA should be approximately 1.85. Deviations from this value are indicating quality deficiencies caused by impurities, such as RNA or protein (Sambrook, 1989).

Agarose gel electrophoresis was performed with 0.8-2% TAE-buffered agarose gels supplemented with 10  $\mu\text{g/ml}$  ethidiumbromide. Size fractionation of DNA fragments was carried out at 80-100 V in TAE buffer. Prior to electrophoresis, loading buffer (6x loading buffer: 60% glycerol, 0.25% bromphenol blue, 0.25% xylene cyanol FF) was added to the DNA samples to a 1x concentration.

Primers were purchased from Metabion (Martinsried, Germany); DNA sequencing was performed by Medigenomix GmbH (Martinsried, Germany) or Sequiserve (Vaterstetten, Germany) or MPI sequencing service.

### **4.3.3 PCR amplification**

DNA was amplified using PCR (polymerase chain reaction) according to the standard protocol mentioned below. Modifications in the reaction setup and the running conditions were made when necessary.

### Typical PCR reaction

DNA template	100-200 ng
Primers	20 pmol each
dNTPs	200 $\mu$ M each
Pfu DNA Polymerase	3 U/50 $\mu$ l
Polymerase buffer	1x

### Typical PCR cycling conditions (30 cycles):

Initial denaturation	95°C for 2 min
Cycle denaturation	95°C for 30-60 s
Annealing	50-58°C for 30-60 s
Extension	72°C for 2-6 min (~40 s /1.0 kb of DNA)
Final extension	72°C for 10 min
Pause	4°C indefinite

#### **4.3.4 DNA restriction digestions and ligations**

DNA restriction digestions were performed according to the product instructions of the respective enzymes. Typically, a 60  $\mu$ l preparative reaction containing the purified PCR product or plasmid DNA: restriction enzyme = 1 $\mu$ g:10U, 0.1 mg/ml BSA (if necessary) in the appropriate reaction buffer was used. In order to avoid religation, dephosphorylation of vector DNA cohesive ends was carried out using calf intestinal alkaline phosphatase (CIP) following subsequent purification with QIAquick PCR purification and gel extraction kit or Wizard SV Gel and PCR Clean-Up System according to the manufacturer's instructions.

DNA ligations were performed in the presence of T4 DNA ligase. Typically, for a 10 µl reaction containing 1 µg DNA consisting of dephosphorylated vector DNA and an insert fragment in a molar ratio between 1:3 and 1:10, 1 µl T4 DNA ligase (400 U/µl) and 1x ligase buffer was used. The ligation was carried out for 1 h at 25°C or overnight at 16°C for increased efficiency. The complete reaction was used to transform chemically competent *E. coli* DH5α cells.

For processing of constructs resulting from whole plasmid PCR (for pont mutations or for inserting FLAG or His6 tags) a subsequent digestion with *DpnI* for 1 h at 37°C was performed to destroy methylated original template DNA. After enzyme inactivation for 20 min at 80°C and DNA-purification, ligation was carried out in 10 µl reactions containing 8 µl DNA, T4 DNA ligase reaction buffer and 400 U of T4 DNA ligase. After incubation for at least 2 h at RT, reactions were transformed into chemically competent *E. coli* DH5α cells.

### **4.3.5 Preparation and transformation of competent *E. coli* cells**

Chemically competent *E. coli* BL21(DE3) or DH5α cells were prepared by the Rubidium Chloride method (Hanahan, 1983). 50 ml LB medium was inoculated (1:100) with an overnight culture of the respective strain and grown at 37°C to an OD<sub>600</sub> of 0.4-0.6. The culture was chilled on ice for 15 min before it was centrifuged for 15 min at 2500 rpm and 4°C. Cells were resuspended in 16 ml buffer I (100 mM RbCl, 50 mM MnCl<sub>2</sub>, 30 mM KOAc, 10 mM CaCl<sub>2</sub>, 15% (v/v) glycerol, pH 5.8 adjusted with acetic acid) and incubated on ice for 15 min. Then, cells were centrifuged again and resuspended in 4 ml buffer II (10 mM MOPS, 10 mM RbCl, 75 mM CaCl<sub>2</sub>, 15 % (v/v) glycerol, pH 6.8 adjusted with NaOH). After incubation on ice for 15 min, aliquots of 50 µl were frozen on dry ice and stored at -80°C.

For transformations, competent cells were thawed on ice, immediately mixed with 1 µl plasmid DNA or 10 µl of a ligation reaction and incubated on ice for 30 min. Cells

were then subjected to a heat shock at 42°C for 45 s and immediately chilled on ice for 2 min. 800 µl of prewarmed LB medium was added and upon phenotypical expression for 60 min at 37°C, the transformation reaction was plated on selective LB-agar plates and incubated at 37°C overnight.

#### 4.3.6 Cloning strategies

*At-rbcL* and *At-rbcS1A* and *At-rbcS3B* were cloned by Karnam Vasudeva Rao. A cDNA-library of *Arabidopsis thaliana* was used to amplify the genes of *At-rbcL* (aa 1-479, AtCg00490), *At-rbcS1A* (aa 56-180, AT1G67090), *At-rbcS3B* (aa 56-181, AT5G38410), and *At-rbcX* (aa 83-203, AT5G19855). The transit peptide for *At-RbcS1A*, *At-RbcS3B* and *At-RbcX* was omitted based on ChloroP 1.1 Server (expasy.org) prediction and by multialigning (MultAlin) (Corpet, 1988) with cyanobacterial *RbcS* and *RbcX* respectively. Since four genes encode four types of *RbcS* in *A. thaliana*, those genes with highest expression levels in planta were chosen for this study (Yoon et al., 2001). The genes were cloned between the *NdeI*/*BamHI* sites of plasmid pET11a (Amp<sup>R</sup>) resulting in *At-rbcL*-pET11a, *At-rbcS*-pET11a and *At-rbcX*-pET11a.

The reverse primer for *At-rbcL* contained an additional *NheI* site between the stop codon and the *BamHI* site, the *At-rbcL* expression cassette was extracted from *At-rbcL*-pET11a with *SphI*/*NheI* and introduced between the *SphI*/*XbaI* sites of *At-rbcX*-pET11a, in front of the ribosomal binding site of *At-rbcX*. In the resulting co-expression plasmid *At-rbcLX*-CoEx-pET11a, both *At-rbcL* and *At-rbcX* share the same T7-promotor for co-transcription, but have their own ribosomal binding sites for translation.

pHUE vector has His6+Ubiquitin and a multiple cloning site (Catanzariti et al., 2004). *At-rbcX* was subcloned from *At-rbcX*-pET11a into pHUE between the *SacI*/*BamHI* sites. The FLAG-tagged construct *At-rbcX*<sub>N-FLAG</sub>-pHUE was obtained by introducing the sequences for FLAG-tag (MDYKDDDDK) and a connecting spacer (AG) after the sequence encoding Ubiquitin (before *At-rbcX*) via whole plasmid PCR (Weiner et al., 1994). Point mutations in *At-rbcX* were introduced by the same method using

mutation site-specific primers. Plasmids that were generated in this study were confirmed by sequencing either by Medigenomix GmbH (Martinsried, Germany) or by the MPI sequencing service.

Comparative sequence analysis was performed using BLAST ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and MultAlin-analysis program (Corpet, 1988) <http://bioinfo.genotoul.fr/multalin/multalin.html>).

The constructs for cyanobacterial RbcX and Rubisco components, At-ch-cpn60, At-ch-cpn20 and At-ch-cpn10 were made by Sandra Saschenbrecker.

#### **4.4 Protein biochemical methods**

##### **4.4.1 Protein analytical methods**

###### **4.4.1.1 Determination of protein concentrations**

Concentrations of purified proteins were determined spectrophotometrically on the basis of the Beer-Lambert law and their theoretical extinction coefficients at  $\lambda=280$  nm (Gill and von Hippel, 1989), as calculated by the ProtParam tool at the ExPASy proteomics server (<http://www.expasy.org>) unless mentioned otherwise. Molar concentrations of proteins present as complexes in their native state (e.g. chaperones or Rubisco) are expressed for the native state oligomers.

Protein concentrations of complex protein mixtures and cell lysates were determined spectrophotometrically at  $\lambda=595$  nm using the Bio-Rad protein assay reagent according to the manufacturer's recommendations (Ausubel, 1997; Bradford, 1976).



#### 4.4.1.2 Sodium-dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed using a discontinuous buffer system under denaturing and reducing conditions (Laemmli, 1970). Typically, gels were poured with a 5% polyacrylamide stacking gel on top of a 8-16% polyacrylamide separating gel, depending on the required resolution (Table i). SDS sample buffer was added to the protein samples to a 1x concentration. Prior to loading, samples were boiled at 95 °C for 5 min. Electrophoresis was carried out in Mini-Protean electrophoresis chambers in running buffer at a constant voltage of 150 V.

Chemicals (For 4 mini-gels)	Separating gel			Stacking gel
	8 %	12.5 %	16 %	5 %
30 % Acrylamide (0.8% bis)	4.5 ml	6.8 ml	9.0 ml	1.66ml
1.875 M Tris-HCl, pH 8.8	3.5 ml	3.5 ml	3.5 ml	-
0.6 M Tris-HCl, pH 6.8	-	-	-	1.00 ml
ddH <sub>2</sub> O	8.7 ml	6.4 ml	4.2 ml	7.20 ml
10 % (w/v) SDS	167 µl	167 µl	167 µl	100 µl
10 % (w/v) APS	100 µl	100 µl	100 µl	50 µl
TEMED	10 µl	10 µl	10 µl	10 µl

**Table i. Composition for SDS-PAGE gels.**

**2x SDS sample buffer:** 14 mM Tris-HCl, pH 6.8, 5% SDS, 20 % (v/v) glycerol, 0.1% (w/v) bromphenolblue, 2% (v/v) β-mercaptoethanol

**Running buffer:** 25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS

#### 4.4.1.3 Native PAGE

In Native PAGE the separation of proteins is based on the charge and the hydrodynamic size. Gels were cast as mentioned in Table ii and samples were taken up in 2x native loading dye (50% (v/v) glycerol, 0.25% (w/v) bromphenolblue, in native electrophoresis buffer). Electrophoresis was performed in Mini-Protean electrophoresis

chambers in native electrophoresis buffer (50 mM Tris, 38 mM glycine, pH not adjusted) at 4°C, employing a constant voltage of 100 V for the first 30 min and 200 V till the end of the run.

Chemicals (For 2 mini-gels)	Separating gel		Stacking gel
	6 %	10 %	5 %
Acrylamide/Bis (37.5:1, 30 %)	1.67 ml	2.83 ml	0.83 ml
1.875 M Tris-HCl, pH 8.8	1.75 ml	1.75 ml	-
0.6 M Tris-HCl, pH 6.8	-	-	0.50 ml
ddH <sub>2</sub> O	5.00 ml	3.84 ml	3.65 ml
10 % (w/v) APS	50 µl	50 µl	25 µl
TEMED	5 µl	5 µl	5 µl

**Table ii. Composition for Native-PAGE gels.**

#### 4.4.1.4 Tricine-PAGE

Tricine gels were used for analyzing small proteins (5-20 kDa). Protein samples were mixed with 3x sample buffer (200 mM Tris-HCl, pH 6.8, 2% SDS, 40% (v/v) glycerol, 0.04% (w/v) Coomassie blue G-250, 2% (v/v) β-mercaptoethanol) and boiled for 5 min at 95 °C, before they were applied to the gels prepared according to Table iii. Gels were run in Mini-Protean electrophoresis chambers with separate Cathode-buffer (100 mM Tris, 100 mM Tricine, 0.1 % (w/v) SDS) and Anode-buffer (200 mM Tris-HCl, pH 8.9) at a constant voltage of ≤ 130 V.

Chemicals (For 2 mini-gels)	Separating gel	Stacking gel
	16 %	4 %
30 % Acrylamide (0.8% bis)	5.35 ml	0.65 ml
3 M Tris-HCl, pH 8.45	3.35 ml	1.24 ml
ddH <sub>2</sub> O	200 µl	3.10 ml
10 % (w/v) SDS	100 µl	50 µl
Glycerol	1.00 ml	-
10 % (w/v) APS	100 µl	50 µl
TEMED	10µl	5 µl

**Table iii. Composition for Tricine-PAGE gels.**

#### 4.4.1.5 Bis-Tris Native PAGE

In the Bis-Tris Native PAGE system, protein migration correlates with molecular masses of proteins and was performed according to described procedures (Hansen et al., 1999; Schagger and von Jagow, 1991). Gels were prepared as mentioned in Table iv. Samples were mixed with loading dye to final concentrations of 50 mM Bis-Tris, pH 7.0, 5% glycerol and 0.1% bromphenolblue. Electrophoresis was performed in cold room or on ice in Mini-Protean electrophoresis chambers. The cathode buffer consisted of 50 mM Tricine and 15 mM Bis-Tris, pH 7.0, whereas the anode buffer contained 50 mM Bis-Tris, pH 7.0. Initially the gels were run at 100 V for 40 minutes and further at 240 V.

Chemicals (For 4 mini-gels)	Separating gel	Stacking gel
	12.5 %	5 %
30 % Acrylamide (0.8% bis)	4.16 ml	2.20 ml
3x gel buffer (150 mM Bis-Tris, pH 7.0, 1.5 M $\epsilon$ -amino n-caproic acid)	3.34 ml	4.00 ml
Glycerol	1.58 ml	-
ddH <sub>2</sub> O	0.88 ml	5.68 ml
10 % (w/v) APS	33.4 $\mu$ l	96 $\mu$ l
TEMED	3.4 $\mu$ l	9.6 $\mu$ l

**Table iv. Composition for Bis-Tris Native PAGE gels.**

#### 4.4.1.6 Coomassie blue staining of polyacrylamide gels

Coomassie blue staining was carried out to detect protein amounts of  $\geq 0.5 \mu\text{g}$  on SDS-PAGEs. To fix and to stain the proteins, gels were incubated in staining solution (0.16% (w/v) Coomassie brilliant blue R-250, 40% (v/v) ethanol, 8% (v/v) acetic acid) followed by multiple washes in destaining solution (20% (v/v) ethanol, 7% (v/v) acetic acid) to remove background staining.

### 4.4.1.7 Silver staining of polyacrylamide gels

After electrophoresis, gels were incubated for 30 min in fixing solution (40% (v/v) ethanol, 10% (v/v) acetic acid), resulting in precipitation of the proteins and diffusion of SDS. Subsequently, gels were placed into incubation solution (300 ml ethanol, 68 g/l sodium acetate x 3 H<sub>2</sub>O, 2 g/l sodium thiosulphate x 5 H<sub>2</sub>O) for 30 min to oxidize the proteins. Gels were then washed in water three times for 5 min and transferred into silver solution (1 g/l silver nitrate, 0.025% (v/v) formaldehyde added before use) for 40 min. Thereafter, proteins were visualized by replacement of the silver solution with developing solution (26 g/l sodium carbonate, 0.0125% (v/v) formaldehyde added before use). The sodium carbonate of the latter solution reduces the silver nitrate attached to the proteins and thus the proteins adopt a brown color. As soon as the desired staining intensity was achieved, the reaction was stopped by addition of stop solution (20% (v/v) ethanol, 7% (v/v) acetic acid), which was replaced by water after 20 min.

### 4.4.1.8 Autoradiography

Autoradiography was used to detect radioactively (<sup>35</sup>S) labeled proteins. Polyacrylamide gels were Coomassie stained, destained, rinsed in water and vacuum dried (Stacked Gel Dryer SGD300) on whatman paper. Migration distances of standard proteins were marked on the whatman paper with a radioactive dye. The radioactive products from the gel was transferred to a phosphoimaging plate using the phosphoimaging cassette (exposure time 3-8 hours) which was read by phosphoimager (Image Reader Fuji-FLA2000) and processed using AIDA software.

### 4.4.1.9 Western blotting and immunodetection

Western blotting was carried out in a semi dry blotting unit. Upon separation by SDS-PAGE, proteins were transferred to nitrocellulose membranes by applying a constant current of ~1 mA/cm<sup>2</sup> gel size in 25 mM Tris, 192 mM glycine, 20% methanol for 1.5 h. Prior to blocking the membranes with 5% skim milk powder in TBS (2.1.3.2) for 1 h,

transfer efficiency was verified by PonceauS (2.1.3.2) staining. The membranes were then incubated with primary antibodies (diluted to a suitable concentration in 5% milk TBS) for 1 h at RT or overnight at 4°C, followed by the incubation with HRP conjugated secondary antibodies (diluted 1:1000 in milk TBS) for 1 h at RT. Extensive washing between the incubation steps was performed with TBST. For immunodetection, ECL chemiluminescence solution was freshly prepared by mixing equal amounts of ECL solution I (100 mM Tris-HCl, pH 8.5, 2.5 mM luminol (3-aminophthalhydrazid), 400  $\mu$ M p-coumaric acid) and ECL solution II (100 mM Tris-HCl, pH 8.5, 5.4 mM H<sub>2</sub>O<sub>2</sub>). Membranes were incubated in the resulting solution and protein bands were detected and documented with the Fuji-LAS3000 luminescence and densitometry system.

#### **4.4.1.10 TCA precipitation**

TCA was added to the protein samples at a final concentration of 20% (v/v). After incubation on ice for 15 min and centrifugation (15 min, 20800 x g, 4°C), 200  $\mu$ l of chilled acetone were added. After another centrifugation, the supernatant was removed and pelleted samples incubated at RT until all residual acetone was evaporated. Pellets were dissolved in 20  $\mu$ l 1 x SDS-loading dye (7 mM Tris-HCl, pH 6.8, 2.5% SDS, 10% (v/v) glycerol, 0.05% (w/v) bromphenolblue, 1% (v/v)  $\beta$ -mercaptoethanol), mixed with 1  $\mu$ l 2 M Tris-base, boiled for 5 min at 95°C and subjected to SDS-PAGE analysis.

#### **4.4.1.11 FFF-MALS (field flow fractionation - multiangle light scattering)**

Mass Determination of proteins by FFF-MALS was performed by Dr. Manajit Hayer-Hartl. Protein complexes (80  $\mu$ g) were analyzed by field flow fractionation (FFF) using a 490 nm spacer and 30 kDa MWCO membrane (Wyatt Technology) with elution and crossflow of 1 ml/min (Roessner, 1994). The FFF was online with DAWN EOS multi-angle light scattering (Wyatt Technology, 690 nm laser), variable wavelength UV absorbance set at 280 nm (Agilent 1100 series) and Optilab DSP refractive index (Wyatt Technology, 690 nm) detectors (Wyatt, 1993). Masses were calculated using the ASTRA

software (Wyatt Technology) with a value set to for  $dn/dc$  for protein of 0.185 ml/g. Alternatively, gel filtration (TSK Super 3000 SW, 4.5 mm x 30 cm) could precede MALS-analysis instead of FFF.

#### 4.4.1.12 N-terminal sequencing of proteins

For N-terminal sequencing, proteins were separated by SDS-PAGE and blotted onto a PVDF membrane, which was soaked in methanol and equilibrated with Western blot buffer prior to protein transfer. The membrane was stained (0.1% (w/v) Coomassie brilliant blue R-250, 10% (v/v) acetic acid, 40% (v/v) methanol) for 2-5 min and subsequently destained (10% (v/v) acetic acid, 30% (v/v) methanol), before it was rinsed in H<sub>2</sub>O and air-dried. Protein bands of interests were cut and analyzed *via* Edman degradation by the *MPI* protein sequencing service.

#### 4.4.1.13 Sequence alignments

Multiple sequence alignment with hierarchical clustering was performed using MultAlin (Corpet, 1988) (<http://bioinfo.genotoul.fr/multalin/multalin.html>).

### 4.4.2 Protein expression and purification

All protein purification steps were performed at 4°C unless stated otherwise. Purification of At-ch-cpn60 $\alpha\beta$ , Syn7002-RbcX, Syn7002-RbcS Syn6301-RbcL<sub>8</sub>S<sub>8</sub> and Syn6301-RbcS were standardized by Sandra Saschenbrecker and the protocol is as described below.

#### 4.4.2.1 At-ch-cpn60 (*Arabidopsis thaliana* ch cpn60 $\alpha_7\beta_7$ )

*E. coli* BL21(DE3), transformed with at-ch-cpn $\alpha\beta$ -CoEx-pET11a, were grown at 37°C in LB medium. After reaching mid-log phase, expression of At-ch-cpn60 $\alpha\beta$  was

induced with 1 mM IPTG for 3.5 h. Cells were harvested by centrifugation (25 min at 4200 rpm), resuspended and incubated for 1 h in buffer A (20 mM Tris-HCl, pH 7.5, 20 mM NaCl), supplemented with 1 mM EDTA, 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail. Cells were disrupted by freeze-thawing as well as ultrasonication and cell debris was removed by ultracentrifugation (35 min at 40000 rpm). The lysate supernatant was applied to a SourceQ column, equilibrated with buffer A/1 mM EDTA and eluted with a linear salt gradient from 0.02 to 1 M NaCl. Throughout the purification, fractions were analyzed both by SDS-PAGE and by Native PAGE to distinguish complexes of At-ch-cpn60 $\alpha_7\beta_7$  from GroEL and from At-ch-cpn60-monomers. Fractions containing At-ch-cpn60 $\alpha\beta$  were pooled, dialyzed against 20 mM Tris-HCl, pH 7.5 and applied to an equilibrated MonoQ ion exchange column (Amersham Biosciences). Elution was performed with a linear gradient from 0 to 0.95 M NaCl and At-ch-cpn $\alpha\beta$  containing fractions were subsequently applied to a Hi-Trap Heparin Sepharose column, equilibrated with 20 mM Tris-HCl, pH 7.5. At-ch-cpn $\alpha\beta$  did not bind to the latter column and was collected in the flow through, which was concentrated using Amicon Ultra MWCO 100 kDa and applied to Superdex 200 gel filtration chromatography in buffer B (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5% (v/v) glycerol). Fractions containing At-ch-cpn60 $\alpha\beta$  were pooled, concentrated (MWCO 100 kDa), flash-frozen in liquid nitrogen and stored at -80°C. Complex concentration was determined spectrophotometrically at 280 nm ( $192150 \text{ M}^{-1}$  complex At-ch-cpn60 $\alpha_7\beta_7$ ). Native PAGE, light scattering and functional assays confirmed purification of active At-ch-cpn60 $\alpha_7\beta_7$ -complexes. ESI-MS and SDS-PAGE (8% resolution) verified the absence of GroEL/GroES and the presence of equal amounts of  $\alpha$ - and  $\beta$ -subunits in the purified complexes.

#### 4.4.2.2 At-ch-cpn20 (*Arabidopsis thaliana* ch-cpn20)

*E. coli* BL21(DE3) cells transformed with at-ch-cpn20-pET11a were grown to mid-log phase at 37°C and induced with 1 mM IPTG for 3 h. Harvested cells were incubated for 1 h in buffer C (50 mM Tris-HCl, pH 7.5, 20 mM NaCl), containing 1 mM EDTA, 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail. After

ultrasonication, cell debris was removed by ultracentrifugation (35 min, 40,000 rpm). The supernatant was applied to a DE52 ion exchange column, equilibrated with buffer C, and eluted with a linear salt gradient from 0.02 to 1 M NaCl. Fractions containing cpn20 or cpn10 were pooled, dialyzed against 50 mM Tris-HCl, pH 8, applied to an equilibrated SourceQ column and eluted with a linear gradient from 0 to 1 M NaCl. The resulting protein pool was concentrated (MWCO 30 kDa) and subjected to Sephacryl S-200 gel filtration chromatography in buffer C. Fractions containing cpn20 or cpn10 were applied to a Hi-Trap Heparin Sepharose column, equilibrated in the same buffer. Cpn20 or Cpn10, eluting in the flow through, were concentrated (MWCO 30 kDa), flash-frozen in liquid nitrogen and stored at -80 °C. Protein concentration was determined spectrophotometrically at 280 nm (30720 M<sup>-1</sup> tetramer cpn20). N-terminal sequencing confirmed that the proteolysis product of At-ch-cpn20, which appeared throughout the purification, was due to C-terminal proteolysis. Light scattering confirmed the tetrameric nature of cpn20.

### 4.4.2.3 At-ch-cpn20<sub>N-His6</sub>

For *E. coli* BL21(DE3) cells harboring plasmid At-ch-cpn20<sub>N-His6</sub>-pProEx, protein expression and cell lysis was performed as described for At-ch-cpn60 with buffer E (20 mM Tris-HCl, pH 7.5, 500 mM NaCl), supplemented with 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail. Soluble cell lysate was applied to a Ni-NTA-agarose column, equilibrated with buffer E. After stepwise washing with buffer E containing 10 mM, 50 mM and 100 mM imidazole, the majority of At-ch-cpn20<sub>N-His6</sub> eluted with 250 mM imidazole. This protein pool was dialyzed against buffer A and applied to a MonoQ column, which was equilibrated with buffer A and developed with a linear salt gradient from 0 to 1 M NaCl. Fractions containing At-ch-cpn20<sub>N-His6</sub> were concentrated (MWCO 10 kDa) and applied to Superdex 200 gel filtration chromatography in buffer A. Fractions containing the desired protein were concentrated, flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined spectrophotometrically at 280 nm.



If the N-terminal His-tag had to be removed, gel filtration was preceded by overnight digestion with TEV-protease (1 mg TEV per 100 mg tagged protein) at 4°C, followed once more by Ni-NTA affinity chromatography with cpn20 eluting in the flow through.

#### **4.4.2.4 Syn6301-RbcL<sub>8</sub>S<sub>8</sub> (*Synechococcus* sp. PCC6301 RbcL<sub>8</sub>S<sub>8</sub>)**

To increase the amount of soluble Syn6301-RbcL<sub>8</sub>S<sub>8</sub> in the *E. coli* lysate, overexpression of GroEL/ES preceded the expression of RbcLS. For this purpose, *E. coli* BL21 (DE3) cells were transformed with Syn6301-*rbcLS*-pET11a and pBAD33ES/EL and grown to OD<sub>600</sub> ~0.6 at 30°C. Then expression of GroEL/ES was induced with 0.4 % (w/v) arabinose for 1.5 h, before cells were shifted to fresh LB medium (w/o arabinose) containing 1 mM IPTG for expression of RbcLS for 3 h at 30 °C. Cells were harvested by centrifugation, incubated for 1 h in buffer C (supplemented with 1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, Complete protease inhibitor cocktail), freeze-thawed and passed through a high pressure cell disruptor. Cell debris was removed by ultracentrifugation and the lysate supernatant was fractionated on a DE52 column with a linear salt gradient from 0.02 to 1 M NaCl in buffer C/1 mM DTT. Fractions were analyzed for Syn6301-RbcL<sub>8</sub>S<sub>8</sub> by SDS-PAGE and immunoblotting against RbcL. The Syn6301-RbcL<sub>8</sub>S<sub>8</sub> pool was supplemented with (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (20% saturation) and applied to a Phenyl-Sepharose column, equilibrated with 50 mM Tris-HCl, pH 7.5, 20% saturation (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 mM DTT. Elution was performed with a linear gradient from 20 to 0% saturation (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Syn6301-RbcL<sub>8</sub>S<sub>8</sub> fractions were dialyzed against 20 mM Imidazol, pH 6.2. After ultracentrifugation, the supernatant was loaded onto an equilibrated MonoQ column and the protein was eluted with a linear gradient from 0 to 0.7 M NaCl. Fractions containing Syn6301-RbcL<sub>8</sub>S<sub>8</sub> were concentrated (MWCO 100 kDa) and passed over a Superose 6 gel filtration column in buffer B. Eluted Syn6301-RbcL<sub>8</sub>S<sub>8</sub> was concentrated (MWCO 100 kDa), snap-frozen in liquid nitrogen and stored at -80°C. Complex concentration was determined spectrophotometrically at 280 nM (705520 M<sup>-1</sup> complex RbcL<sub>8</sub>S<sub>8</sub>). Light scattering, Native PAGE and carboxylation activity confirmed the complex nature and activity of the purified Syn6301-RbcL<sub>8</sub>S<sub>8</sub>.

#### 4.4.2.5 Syn6301-RbcL<sub>8</sub> (*Synechococcus* sp. PCC6301 RbcL<sub>8</sub>)

*E. coli* BL21(DE3) cells, harboring plasmids Syn6301-*rbcL*-pET11a and pG-KJE8, were grown to mid-log phase at 30 °C. Expression of DnaK/DnaJ/GrpE was induced with 0.4 % (w/v) arabinose for 2 h, before cells were shifted for ~3 h to fresh LB medium (w/o arabinose) supplemented with 1 mM IPTG as inducer for Syn6301-RbcL expression. Note that since purification of RbcL<sub>8</sub> was facilitated in the absence of elevated GroEL levels, coexpression of DnaK/DnaJ/GrpE (instead of GroEL/ES) was carried out to increase the amount of soluble RbcL in *E. coli*. As described for At-ch-cpn60, cells were lysed and fractionated in 25 mM Tris-HCl, pH 8, 1 mM EDTA, 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail. The lysate supernatant was applied to a SourceQ ion exchange column, equilibrated with buffer F (50 mM Tris-HCl, pH 8, 50 mM NaHCO<sub>3</sub>, 10 mM MgCl<sub>2</sub>) / 1 mM EDTA, 0.5 mM DTT. Proteins were eluted with a linear salt gradient from 0 to 1 M NaCl. Fractions were analyzed for the presence of RbcL<sub>8</sub> by SDS-PAGE and Native PAGE, followed by immunoblotting, as well as by measurement of carboxylation activity upon addition of RbcS. Fractions with highest activity and most enriched RbcL<sub>8</sub> were pooled and dialyzed against 20 mM Imidazol, pH 6.5, 50 mM NaHCO<sub>3</sub>, 10 mM MgCl<sub>2</sub>, resulting in a white precipitate, which was pelleted and dissolved in 50 mM Tris-HCl, pH 8. After filtration (0.22 µm) and dialysis against buffer F, the protein solution was applied to a MonoQ column and eluted with a linear salt gradient from 0 to 0.7 M NaCl in buffer F. The Syn6301-RbcL<sub>8</sub> containing fractions were concentrated (MWCO 100 kDa) and subjected to Superdex 200 gel filtration chromatography in buffer F. Fractions containing Syn6301-RbcL<sub>8</sub> were pooled, complemented with 10% (v/v) glycerol, concentrated (MWCO 100 kDa), flash-frozen in liquid nitrogen and stored at -80°C. Complex concentration was determined spectrophotometrically at 280 nM (553040 M<sup>-1</sup> complex RbcL<sub>8</sub>). Light scattering and Native PAGE confirmed that the purified protein was assembled Syn6301-RbcL<sub>8</sub>, which showed carboxylation activity upon addition of RbcS.

#### **4.4.2.6 Syn6301-RbcS and Syn7002-RbcS<sub>FLAG</sub> (*Synechococcus* sp. PCC6301 RbcS and *Synechococcus* sp. PCC7002 RbcS<sub>FLAG</sub> from inclusion bodies)**

RbcS was purified from inclusion bodies by modification of previously described methods (Coligan, 2000; Somerville, 1986). *E. coli* BL21(DE3) cells, transformed with Syn6301-*rbcS*-pET11a or Syn7002-*rbcS*<sub>FLAG</sub>-pET11a, were grown to mid-log phase at 37 °C, before induction of RbcS expression with 1 mM IPTG for 3.5 h. The majority of RbcS was found in inclusion bodies. Harvested cells were incubated for 1 h in lysis buffer (40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA, 5 % (v/v) Triton X-100, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, Complete protease inhibitor cocktail). After freeze-thawing, ultrasonication and centrifugation, the pellet was resuspended and washed in 40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA, 5 % (v/v) Triton X-100, 2 M urea. Subsequent to centrifugation, washing of pellets was repeated in 40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA. Pellets were finally dissolved in denaturation buffer (50 mM Tris-HCl, pH 7.5, 6 M GdnHCl, 1 mM EDTA, 5 mM DTT). The denatured RbcS was refolded by dialysis against 50 mM Tris-HCl, pH 8, 1 mM EDTA, 0.1 mM GSH, 0.01 mM GSSG at a concentration of ca. 0.5 mg/ml and finally frozen in liquid Nitrogen for storage at -80°C. The recovery of refolded protein was ca. 65 to 85 %. Protein concentration was determined spectrophotometrically at 280 nM (19060 M<sup>-1</sup> monomer Syn6301-RbcS, 24410 M<sup>-1</sup> monomer Syn7002-RbcS<sub>FLAG</sub>).

#### **4.4.2.7 Syn7002-RbcX (*Synechococcus* sp. PC7002 wild-type/mutant/FLAG-tagged RbcX)**

*E. coli* BL21(DE3) cells, transformed with the respective RbcX plasmids, were used for expression of RbcX upon incuction with 1 mM IPTG for 3.5 h at at 37 °C. Harvested cells were incubated for 1 h in lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM DTT, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, Complete protease inhibitor cocktail) and disrupted by ultrasonication. After removal of cell debris by centrifugation, the supernatant was applied to a SourceQ column, equilibrated with 50

mM Tris-HCl, pH 8.0, 0.5 mM DTT, and eluted with a linear gradient from 0 to 1 M NaCl. Fractions containing RbcX were dialyzed against 20 mM imidazole, pH 6.4, 20 mM NaCl, 0.5 mM DTT, applied to an equilibrated MonoQ column, and eluted with a linear salt gradient up to 0.7 M NaCl. Fractions containing RbcX were dialyzed against buffer G (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub>) and applied to a HiTrap Heparin Sepharose column. RbcX eluted in the flow through, which was then concentrated (MWCO 10 kDa) and subjected to Superdex 200 gel filtration chromatography in buffer G. Fractions containing RbcX were supplemented with 10 % glycerol, concentrated (MWCO 10 kDa), flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined spectrophotometrically at 280 nm (19060 M<sup>-1</sup> dimer Syn7002-RbcX). Syn7002RbcX wild type and mutant proteins were purified in cooperation with Karnam Vasudeva Rao.

#### 4.4.2.8 At-RbcX<sub>N-His6</sub> (*Arabidopsis thaliana* RbcX<sub>N-His6</sub>)

*E. coli* BL21(DE3) cells transformed with At-*rbcX*<sub>N-His6</sub>-pET11a plasmid, were used for expression of At-*rbcX* upon induction with 1 mM IPTG for 7 h at 18°C. Harvested cells were incubated for 1 h in buffer E (20 mM BisTris-HCl, pH 6, 500 mM NaCl, 10mM Imidazole), supplemented with 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail and disrupted by ultrasonication. Soluble cell lysate was applied to a Ni-NTA-agarose column, equilibrated with buffer E. After stepwise washing with buffer E containing 10 mM, 50 mM and 100 mM imidazole, the majority of At-RbcX<sub>N-His6</sub> eluted with 250 mM imidazole. This protein pool was dialyzed against buffer containing 20mM BisTris-HCl, pH6, 100mM NaCl, 8% Glycerol. The protein was concentrated in centricons (MWCO 10 kDa). Concentration was determined spectrophotometrically at 280 nm and the protein was flash-frozen in liquid nitrogen and stored at -80°C.

#### **4.4.2.9 At-RbcX<sub>N-His6+Ub</sub> (*Arabidopsis thaliana* RbcX<sub>N-His6+ Ubiquitin</sub>) and At-RbcX<sub>N-His6+Ub+FLAG</sub> (*Arabidopsis thaliana* RbcX<sub>N-His6+ Ubiquitin+FLAG</sub>)**

*E. coli* BL21(DE3) cells, transformed with either At-*rbcX*-pHUE or At-*rbcX*<sub>N-FLAG</sub>-pHUE plasmids, were used for expression of At-*rbcX* upon induction with 1 mM IPTG for 5 h at 30°C. Harvested cells were resuspended and incubated for 1 h in buffer E (20 mM BisTris-HCl, pH 6, 500 mM NaCl, 10mM Imidazole), supplemented with 0.5 mg/ml lysozyme, 10 U/ml Benzonase and Complete protease inhibitor cocktail and disrupted by ultrasonication. Further the purification was performed using the modification of the published protocol (Catanzariti et al., 2004). Soluble cell lysate was applied to a Ni-NTA-agarose column, equilibrated with buffer E. After stepwise washing with buffer E containing 10 mM, 50 mM and 100 mM imidazole, the majority of His+Ubiquitin tagged At-RbcX eluted with 250 mM imidazole. The fractions containing At-RbcX were pooled and dialyzed against buffer containing 20mM BisTris-HCl, pH6, 100mM NaCl, 8% Glycerol for 8 hours. After dialysis, the protein was concentrated (MWCO 10 kDa), flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined spectrophotometrically at 280 nm.

To remove the N-terminal His+Ub-tag, the samples from the first Nickel NTA containing At-RbcX was incubated with the ubiquitin protease, Usp2-cc at 1:100 protease to protein molar ratio at 16°C, overnight. 3% Glycerol and 1.5M mercaptoethanol were added to the cleavage reaction. After incubation this mixture was dialysed against buffer containing 20mM BisTris-HCl, pH9.2, 100mM NaCl, 10% Glycerol and then subjected to Ni-NTA chromatography. All the cleaved ubiquitin and the protease which are His-tagged bound to the Ni-NTA beads and At-RbcX was in the flow-through. The flow-through was concentrated in a 10kDa cut off centricon and applied to an equilibrated MonoQ ion exchange column (Amersham Biosciences). Elution was performed with a linear gradient from 0 to 1M NaCl and At-RbcX containing fractions were subsequently pooled and concentrated. The concentrated samples were flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined spectrophotometrically at 280 nm.

#### **4.4.2.10 At-RbcS1A (*Arabidopsis thaliana* RbcS1A) from inclusion bodies**

At-RbcS1A was purified from inclusion bodies. *E. coli* BL21(DE3) cells, transformed with At-*rbcS1A*-pET11a, were grown to mid-log phase at 37°C, before induction of *rbcS* expression with 1 mM IPTG for 3.5 h. The majority of RbcS was found in inclusion bodies. Harvested cells were incubated for 1 h in lysis buffer (40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA, 5% (v/v) Triton X-100, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, Complete protease inhibitor cocktail). After freeze-thawing, ultrasonication and centrifugation, the pellet was resuspended and washed in 40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA, 5% (v/v) Triton X-100, 2 M urea. Subsequent to centrifugation, washing of pellets was repeated in 40 mM Tris-HCl, pH 8, 0.25 M sucrose, 10 mM EDTA. Pellets were finally dissolved in denaturation buffer (50 mM Tris-HCl, pH 7.5, 6 M GdnHCl, 1 mM EDTA, 5 mM DTT). The denatured RbcS was refolded by dialysis against 50 mM Tris-HCl, pH 8, 1 mM EDTA, 0.1 mM GSH, 0.01 mM GSSG at a concentration of ca. 0.5 mg/ml and finally frozen in liquid Nitrogen for storage at -80°C. Protein concentration was determined spectrophotometrically at 280 nM.

#### **4.4.3 Functional analyses**

##### **4.4.3.1 ATPase activity assay**

The ATPase activity of chaperonins was determined as described (Lanzetta et al., 1979). Chaperonins were diluted at a concentration of 0.5 µM into assay-buffer 1 (20 mM MOPS-KOH, pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>). If the influence of co-chaperones was analyzed, the latter were added at a final concentration of 1 µM. The reactions were incubated for 5 min at 25°C, before they were initiated by addition of 2 mM ATP. At indicated time points aliquots of 10 µl were withdrawn and the reaction stopped with CDTA (final concentration 20 mM). The resulting samples were mixed with 300 µl color reagent (filtered 3:1 mixture of 0.045% (w/v) malachite green hydrochloride in H<sub>2</sub>O and 4.2% (w/v) ammonium molybdate in 4 N HCl; supplemented before use with 0.1% (v/v)

Triton X-100) and 40  $\mu$ l 37% citric acid. After incubation for 30 min at 25°C, the absorption at 640 nm was measured.

For quantification of ATP hydrolysis, a standard assay was performed, in which 10  $\mu$ l of solutions containing 0, 10, 25, 50, 100, 250, 500 and 1000  $\mu$ M  $K_2HPO_4$  were applied to the colorimetric assay described above. The resulting calibration curve displayed the relation between phosphate concentration and the respective absorbance and could thus be used to determine the ATP hydrolysis rate of the chaperonins.

#### **4.4.3.2 *In vivo* co-expression of RbcL or RbcLS with chaperones in *E. coli* and Rubisco carboxylation activity**

*E. coli* BL21(DE3) cells were transformed with pET-vectors for expression of RbcL, RbcL/X, RbcL/S or RbcL/X/S. If necessary, the resulting strains were additionally co-transformed with pG-KJE8 for GroEL/GroES or/and DnaK/DnaJ/GrpE expression. Single colonies were grown to mid-log phase at 30°C. Expression of the *rbc*-genes from T7-promoters was induced with 1 mM IPTG for ~3.5 h at 30°C with or without prior transient overexpression of GroEL/GroES (induced with 0.4% (w/v) arabinose) or GroEL/GroES+DnaK/DnaJ/GrpE (induced with 0.4% (w/v) arabinose and 20  $\mu$ g/ml tetracycline) for 2 h at 30°C, followed by a transfer to fresh medium.

Equivalent amounts of cells were pelleted, incubated in lysis buffer (50 mM Tris-HCl, pH 8, 20 mM NaCl, 5 mM  $MgCl_2$ , 1 mM EDTA, 0.1% (v/v) Triton X-100, 0.1 mg/ml lysozyme) on ice for 30 min, ultrasonicated and fractionated into soluble and insoluble fractions by centrifugation (20800 x g, 30 min at 4°C). Comparative analysis of total, soluble and insoluble protein was performed by SDS-PAGE. Soluble lysate fractions were analyzed for assembled RbcL by 6% Native PAGE, followed by immunoblotting against RbcL.



Carboxylation activity was determined by diluting aliquots of lysate supernatants into assay-buffer 3 (100 mM Tris-HCl, pH 7.5, 10 mM KCl, 2 mM Mg(OAc)<sub>2</sub>) containing 1 mM DTT and 2 μM BSA. Samples containing only RbcL were supplemented with 7 μM Syn6301-RbcS or Syn7002-RbcS<sub>FLAG</sub> or At-RbcS1A and assembly was allowed to proceed for 5 min at RT. Thereafter, a <sup>14</sup>C-Mix (in 100 mM Tris-HCl, pH 7.5, 10 mM KCl) was added to give final concentrations of 60 mM NaHCO<sub>3</sub>, 0.5 μCi NaH<sup>14</sup>CO<sub>3</sub> and 10 mM MgCl<sub>2</sub>. After incubation for 5 min, carboxylation was initiated by addition of 2.5 mM RuBP and stopped with acetic acid (3 N) after 30 min. The resulting mixes were heated (96°C) until complete evaporation of liquid, the remaining non-volatile components were dissolved in 100 μl water, taken up in 1 ml scintillation fluid and radioactivity of the fixed carbon was quantified (LS 6500 multi-purpose scintillation counter) (Dickson et al., 2000; Goloubinoff et al., 1989a; Goloubinoff et al., 1989b; Viitanen et al., 1995).

#### **4.4.3.3 *in vitro* translation of Rubisco, immunodepletion of GroEL from RTS *E. coli* lysate and Pulse-chase Assays**

T7 promoter constituted plasmid (200-300 ng/μl) for Rr-*rbcL* or Syn6301-*rbcL* or Syn7002-*rbcL* or At-*rbcL* was translated *in-vitro* in the reconstituted *E. coli* lysate (coupled RTS100 *E. coli* HY transcription/ translation system from a bacterial S30 lysate) in the presence of 0.5 U/μl RNAGuard ribonuclease inhibitor, Complete protease inhibitor cocktail (Stock: 1 Complete EDTA free mini tablet/1ml of RTS reconstitution buffer; Final amount used 2 μl/20 μl reaction volume), 0.37 MBq <sup>35</sup>S-methionine per 20μl reaction, 50 μM unlabelled methionine. Wherever indicated necessary chaperones and other proteins were added to the reaction in the concentrations mentioned in the figure legends. Translation was carried out at 30°C for 90 minutes and stopped by addition of Chloramphenicol (CAM, 200 μg/ml) on ice. Post-translational addition of proteins was performed after CAM addition, followed by transfer of reactions back to 30°C.

When indicated, immunodepletion of GroEL from the lysate was performed by incubation with polyclonal GroEL antibody bound to Protein A sepharose beads by gentle



shaking for 45 min at 4°C and removal of beads by centrifugation. Successful depletion of GroEL was confirmed by SDS-PAGE followed by immunoblotting.

RTS products were separated into soluble and insoluble fractions by centrifugation (20800 x g for 30 min at 4°C). The former was analyzed by native-PAGE and SDS-PAGE and the latter by SDS-PAGE followed by autoradiography.

In pulse-chase experiments, translation was performed in presence of 0.5 µM GroEL, 1 µM GroES and 40 µM RbcX proteins (as indicated) for 6 min at 30°C, before <sup>35</sup>S-methionine was added. After 6 min, reactions were chased by addition of unlabelled methionine. Samples were collected at indicated time points and reactions stopped by addition of CAM on ice as described above. Soluble and insoluble fractions were analyzed by discontinuous Bis-Tris Native PAGE (13% resolution gel, 6% stacking gel) or 12.5% SDS-PAGE, respectively, followed by autoradiography.

The *in vitro* translation experiments were performed by Karnam Vasudeva Rao wherever indicated.

#### **4.4.3.4 Analytical gel filtration of *E. coli* lysate or protein complexes**

Gel filtration chromatography of soluble *E. coli* lysate or protein complexes (resulting from co-immunoprecipitation) performed using a Superdex 200 (10/30) gel filtration column, equilibrated in 50 mM Tris-HCl, pH 8, 50 mM NaCl, 5 mM MgCl<sub>2</sub>. If necessary, the sample volumes were reduced (MWCO 30 kDa) to 200 µl prior to application. The column was eluted at 250 µl/min and fractions of 250 µl were collected. If necessary, fractions were TCA-precipitated. Samples were analyzed for RbcL, RbcX<sub>FLAG</sub> or GroEL by SDS-PAGE and Coomassie staining or immunoblotting.

Similarly for the analysis of purified proteins, the protein samples were injected into Superdex 200 (3.2/30) column equilibrated in buffer 20 mM Tris-HCl, pH 9.2, 100 mM NaCl and 10% Glycerol.

#### 4.4.3.5 Tryptophan-fluorescence spectroscopy

Tryptophan-fluorescence was measured using Spex Fluorolog 3 with the following parameters: excitation 295 nm, emission scan 315-450 nm, incr. 2 nm, interval time 1 sec, slits 1/5. In order to monitor tryptophan-fluorescence of RbcL in refolding reactions, Rr-RbcL or Syn6301-RbcL<sub>8</sub> was denatured in denaturation-buffer (20mM MOPS-KOH, pH 7.5, 6 M GdnHCl, 10 mM KCl, 1 mM EDTA, 10 mM DTT) 1 for 60 min at 25 °C. The denatured protein was diluted 100-fold (0.25 µM RbcL monomer) into ice-cold assay-buffer (20 mM MOPS-KOH, pH 7.5, 100 mM KCl, 5 mM Mg(OAc)<sub>2</sub>) containing 0.5 µM GroEL. After incubation for 10 min on ice, a tryptophan fluorescence scan was taken. 1 µM GroES was added to this sample and a tryptophan fluorescence scan was taken again. Refolding was started by supplementing the reaction with 2 mM ATP and the kinetics of refolding was measured by monitoring the change in tryptophan fluorescence over time. At the end of kinetics run (after ~35 min), a tryptophan scan was performed. If only native or denatured substrate or merely substrate-binding to GroEL had to be analyzed, reactions were modified accordingly. Background fluorescence of chemically identical reactions lacking Rubisco was subtracted in each assay.

#### 4.4.3.6 ANS-fluorescence spectroscopy

Analysis of ANS (1-anilino-8-naphthalene-sulphonate) fluorescence was performed by the same method as described for tryptophan fluorescence with the exception that assay-buffer 1 contained 1 µM ANS. Fluorescence was measured with Spex Fluorolog 3 applying the following parameters: excitation 350 nm, emission scan 410-570 nm, incr. 2 nm, slits 1/5. Single ring version of GroEL was used for the refolding reactions and low salt buffer containing 20 mM MOPs/KOH pH 7.5, 10 mM KCl, 5 mM Mg(OAc)<sub>2</sub> was used. Data were corrected for background fluorescence.

#### **4.4.3.7 Circular Dichroism Spectroscopy**

Far-UV CD spectra and melting curves were measured with a Jasco J-715 spectrometer equipped with a Peltier-thermostat using 0.1 cm cuvette. Wavelength scans of Syn7002-RbcX wild type and mutant proteins (Q29A, Y17AY20L, R70A, P87A, R102A), Syn7002-RbcX<sub>NFLAG</sub>, At-RbcX<sub>N-FLAG</sub> proteins were recorded at 30°C at a protein concentration of 0.1mg/ml. Syn7002-RbcX wild type and mutant proteins were analyzed in a buffer containing 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub>. For At-RbcX<sub>N-FLAG</sub>, 20mM BisTris-HCl, pH 9.2, 100mM NaCl was used.

#### **4.4.4 Crystallography and structure analysis**

##### **4.4.4.1 Analytical subtilisin digestion of Syn7002-RbcX**

Syn7002-RbcX was digested at a concentration of 1 mg/ml by different concentrations of subtilisin (0-1000 µg/ml) in 20 mM HEPES-NaOH, pH 7.5, 50 mM NaCl on ice. Samples were withdrawn after 60 min, respectively, and supplemented with PMSF at a final concentration of 10 mM to stop the digest. Samples were analyzed on 16% Tricine-PAGE. Truncated Syn7002-RbcX was identified as Syn7002-RbcX (ΔC25) by Edman degradation (N-terminal sequencing) and ESI-MS (Saschenbrecker et al., 2007).

##### **4.4.4.2 Protein crystallization**

##### **Syn7002-RbcX mutants**

Syn7002-RbcX mutant proteins were purified in cooperation with Karnam Vasudeva Rao. Crystals were grown using the hanging drop vapor diffusion method at 20 °C by mixing 1 µl protein sample and 1 µl reservoir solution. The conditions were chosen based on Syn7002RbcX wildtype crystallization trials. Syn7002-RbcX (Q29A) crystallized in 100mM HEPES-NaOH, pH 7.5, 1.5 M Sodium acetate. Syn7002-RbcX (Y17A,Y20L) crystallized in 100mM HEPES-NaOH, pH 7.5, 3 M Sodium acetate and Syn7002-RbcX(R70A) crystals were obtained in 100mM HEPES-NaOH, pH 7.5, 2.4 M Sodium

acetate. For cryo-protection, the crystals were transferred stepwise into mother liquor containing 30 % (v/v) glycerol.

### **Syn7002-RbcX / Syn7002-RbcL peptide**

Crystallization of Syn7002-RbcX/peptide complex was performed by Sandra Saschenbrecker (Saschenbrecker et al., 2007). Crystals of Syn7002-RbcX were incubated for 1 h in buffer (0.1 M HEPES-NaOH, pH 7.5, 1.4 M sodium acetate) containing 1 mM Syn7002-RbcL peptide, EIKFEFD. For cryo-protection, the crystals were transferred stepwise into mother liquor containing 1 mM peptide and 20 % (v/v) glycerol. For co-crystallization trials, a mixture of Syn7002-RbcX (in 10 mM Tris-HCl, pH 7.5) and Syn7002-RbcL peptide (EIKFEFD (in H<sub>2</sub>O) at a molar ratio of 1:1.5 (467  $\mu$ M RbcX-dimer, 700.5  $\mu$ M peptide) was used and crystallization performed as above. Well-diffracting RbcX crystals of space group *I*222 were obtained from PEG / Ion Screen condition 14 (0.2 M potassium thiocyanate, 20% (w/v) polyethylene glycol 3350) (Hampton Research, USA). These crystals apparently contained no bound peptide.

#### **4.4.4.3 Structure determination**

Structure determination of RbcX mutants was performed in cooperation with Dr. Andreas Bracher and is as follows.

Diffraction data were collected at beamlines mentioned in Table1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Diffraction data were integrated with Mosflm (Leslie, 1992) and Scala (Evans, 1997).

The structure of Syn7002-RbcX wild type was solved by SIRAS using crystals incubated with 0.2 mM K<sub>2</sub>(PtCl<sub>4</sub>) in mother liquor for 22 h. Six Platinum sites were found by Patterson methods using Solve (Terwilliger and Berendzen, 1999). Density modification was carried out with Resolve (Terwilliger, 2000). The resulting map was

readily interpretable, and continuous backbone density for all six RbcX chains between residues 4 and 109 could be traced with O (Jones et al., 1991). Subsequent iterative model improvement and refinement were performed with O and Refmac5 (Murshudov et al., 1997). The high resolution crystal form of Syn7002-RbcX was solved by molecular replacement using Molrep (Vagin and Isupov, 2001). This structure and the RbcX mutant structures were edited with Coot (Emsley and Cowtan, 2004) and refined with Refmac. In all structures, residues facing solvent channels without detectable side chain density were modeled as alanines. All models obey good stereochemistry with no Ramachandran outliers as judged by the program Procheck (Laskowski, 1993). Coordinates were aligned with Lsqman (Kleywegt, 1994). The similarity score was calculated with ESPript using the Rissler scoring matrix (Gouet et al., 1999). Figures were generated with the programs Molscrip (Kraulis, 1991), Bobscript (Esnouf, 1997), Raster-3D (Meritt and Bacon, 1997), and Pymol 11 (DeLano, W.L. The PyMOL Molecular Graphics System (2002) on World Wide Web (<http://www.pymol.org>)).

### **4.4.5 Electron microscopy**

Electron microscopic analyses were done in collaboration with the Department of Molecular Structural Biology, MPI, Martinsried.

#### **4.4.5.1 Preparation of At-ch-cpn60/At-ch-cpn20 complexes for Electron microscopy**

At-ch-cpn60 (1  $\mu$ M final concentration), At-ch-cpn20 (0.9  $\mu$ M) and 2mM ATP were incubated in 50 mM HEPES, pH 7.5, 50mM MgCl<sub>2</sub>, 50mM KCl, at 25<sup>0</sup>C for 5 min and used immediately for the preparation of grids.

#### **4.4.5.2 Preparation of Negatively stained specimens**

Negative staining was performed using the single droplet procedure (Harris, 1991). Carbon support films were briefly glow-discharged (30 sec) in a partial atmosphere to

render them hydrophilic and adsorptive. At-ch-cpn60 $\alpha\beta$ , At-ch-cpn60 $\alpha\beta$ /At-ch-cpn20 complexes were diluted to 0.1 mg/ml with 20mM Tris-HCl (pH 7), 50mM NaCl, 5% Glycerol and applied to the carbon films. After 30 sec most of the protein solution was removed with a filter paper and replaced by a droplet of 2% uranyl acetate. After similarly removing the excess uranyl acetate, grids were dried at room temperature and observed under the transmission electron microscope.

#### **4.4.5.3 Recording of micrographs**

Transmission electron microscopy was performed using the Philips CM12. Electron micrographs were recorded at 120kV and at instrumental magnifications of 22,000 or 28,000 as mentioned.

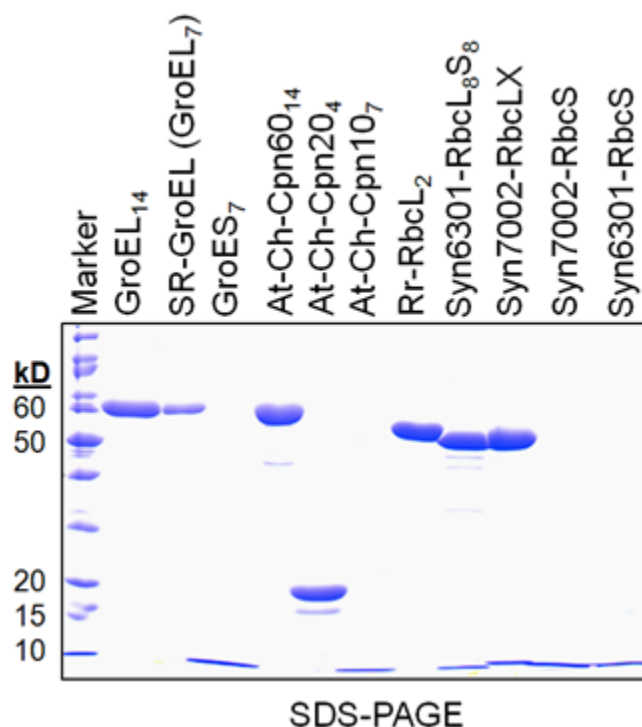
#### **4.4.5.4 Single Particle Analysis**

The image processing steps were performed using the EM software package (Hegerl, 1996; Hegerl and Altbauer, 1982). All particles recognizable as chaperonin side views and top views, not in contact with other particles were manually selected from the micrographs. These images were first subjected to 10 rounds of alignment and classification, specifying 20 output classes. A reference-free alignment, multivariate statistical data compression and automatic classification were used to obtain class averages. Unique averages were selected from the resulting class averages.

## 5 Results

Evidence for chaperonin-mediated Rubisco folding in the chloroplasts was found over 2 decades ago when studies revealed that newly synthesized large subunits were transiently associated with a large multimeric complex prior to appearance in the holoenzyme through an ATP-dependent reaction (Ellis, 1991; Milos, 1984). It is now well documented that for the folding of higher plant Rubisco, chloroplast homologs of the GroEL and GroES chaperones are necessary (Gatenby and Ellis, 1990; Goloubinoff et al., 1989a; Gutteridge and Gatenby, 1995). The absence of an *in vitro* or bacterial expression system for the assembly of higher-plant hexadecameric Rubisco suggests a unique nature for the chloroplast chaperonins and is testimony to the limitations of our current understanding. There may be other requirements for the successful folding and assembly of hexadecameric Rubisco (Houtz and Portis, 2003). However, the Form II Rubisco from *Rhodospirillum rubrum* (Rr-RbcL<sub>2</sub>) can be folded and assembled as active enzyme *in vivo* in *E. coli* and *in vitro* upon interaction with GroEL/ES (Brinker et al., 2001; Goloubinoff et al., 1989a) .

Numerous electron microscopic and crystallographic insights into the structure of GroEL are available (Langer et al., 1992b; Saibil et al., 1991; Hartl and Hayer-Hartl, 2002). GroEL is a tetradecameric complex consisting of identical 57 kDa subunits. The electron microscopic studies have revealed that GroEL has a double toroidal structure with the central cavity of ~50 Å (Hartl, 1996). However, the structure for chloroplast cpn60 is not well established.



**Figure 19. Proteins used in this study**

(A) Coomassie stained SDS-PAGE of recombinantly purified *E. coli* GroEL, Single-Ring GroEL (SR-GroEL) and GroES, *A. thaliana* ch-pn60, ch-cpn20 and ch-cpn10, *Rhodospirillum rubrum* RbcL<sub>2</sub>, *Synechococcus* sp. PCC7002 RbcS, *Synechococcus* sp. PCC6301 RbcS.

## 5.1 Structural characterization of chloroplast cpn60 and cpn20 from *Arabidopsis thaliana*

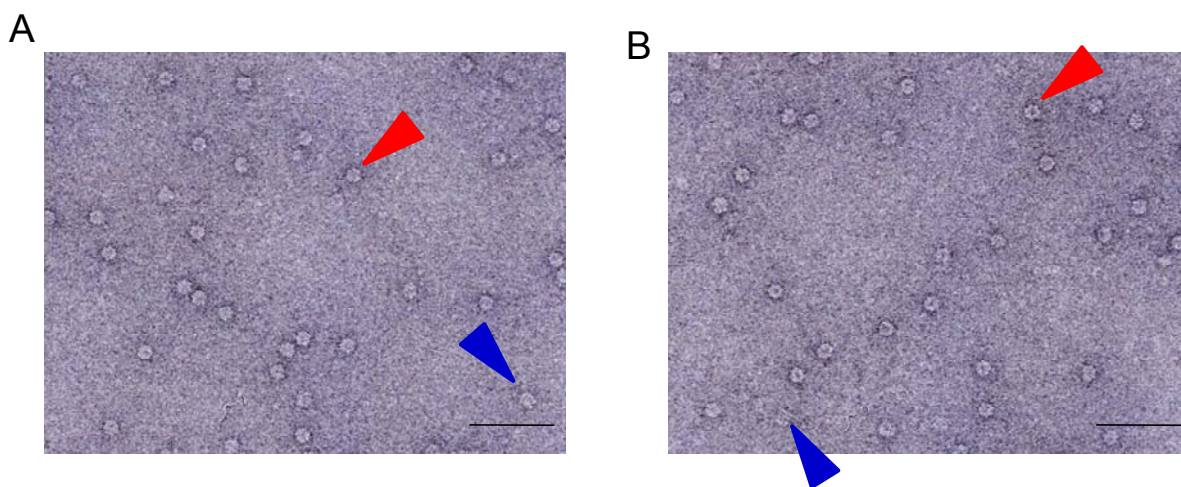
### 5.1.1 Electron microscopic analysis of *A. thaliana* cpn60

In order to know the general structure of *A. thaliana* chloroplast cpn60 (At-ch-cpn60), the recombinantly purified protein (Figure 19, lane 5) was subjected to electron microscopic analysis. The protein was stained with uranyl acetate and the electron micrographs were recorded with Phillips CM12 microscope at an instrumental magnification of 22,000. Two distinct types of single particle projections were observed in the micrographs, first type resembling a ‘seven-pointed’ star (top-view or end-on view) (Figure 20A and 20B, red triangles) and the second resembling a rectangle (side-view or



## RESULTS

side-on view) (Figure 20A and 20B, blue triangles). From the micrographs, 2543 top views and 477 side views were manually selected. The images were aligned with respect to translation and orientation using standard correlation techniques; the alignment was refined iteratively. In order to detect inter-image structural variations the aligned images were subjected to a classification procedure based on eigenvector-eigenvalue data analysis (van Heel and Frank, 1981).



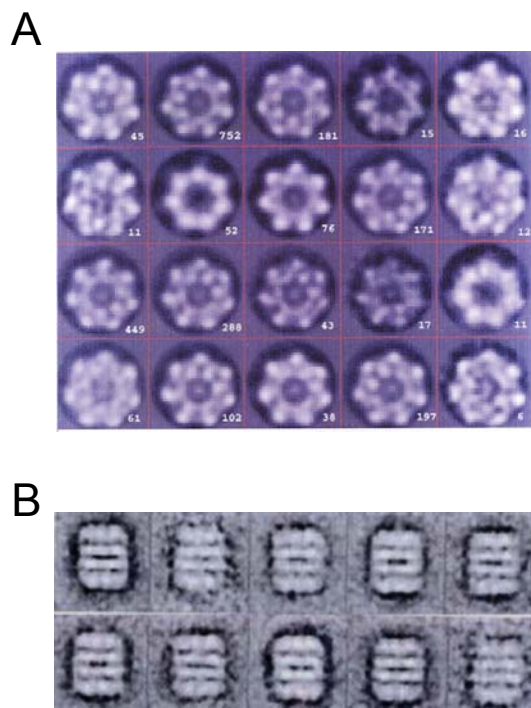
**Figure 20. Electron microscopic analysis of At-ch-cpn60**

(A) and (B) Negatively stained images of At-ch-cpn60. Electron micrographs were recorded at an instrumental magnification of 22,000. The scale bar represents 100nm. Red triangles indicate the particles in the end-on orientation and the blue triangles indicate the particles in the side-on orientation.

During the analysis of the top-view a 7-fold rotational symmetry was imposed to each particle (Figure 21A). The final image (7-fold rotational symmetrized) has a clear visible handedness with seven elongated stain-excluding regions (Figure 22A). Averages obtained from the major classes of top views revealed that the symmetrized diameter of the ch-cpn60 was ~15.3 nm, which is very close to the value reported for *E. coli* GroEL (diameter 15.6 nm, Langer et al., 1992b). The major class averages obtained during the analysis of side views (Figure 21B) and the final image of these appear rectangular with four stripes parallel to the short side (Figure 22B). The dimension of ch-cpn60 rectangle was 16.5 x 14.8 nm which is close to the dimension of GroEL (15.5 x 14.1 nm, (Langer et al., 1992b)). From these observations, the tetradecameric nature of ch-cpn60 consisting

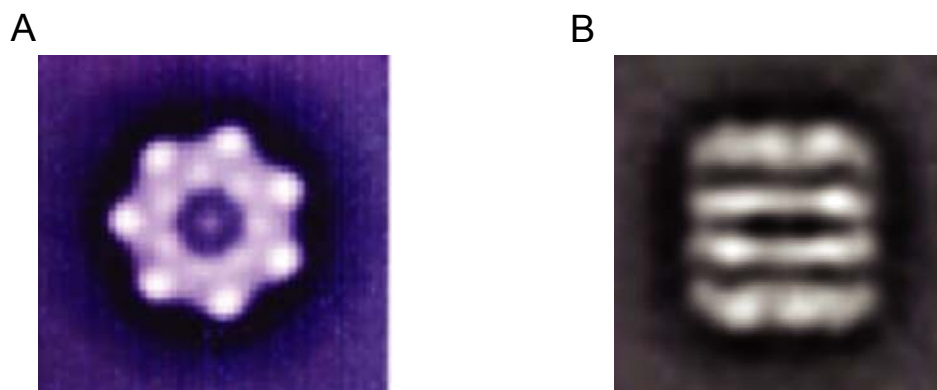
## RESULTS

of two heptameric rings stacked back to back was evident, similar to pea ch-cpn60 (Tsuprun, 1991).



**Figure 21. The major structural classes obtained during averaging of At-ch-cpn60**

The pictures represent major structural classes of end-on (A) and side-on views (B) of At-ch-cpn60 respectively obtained by eigenvector-eigenvalue data analysis. Number in the corner depicts the number of particles belonging to the respective class.



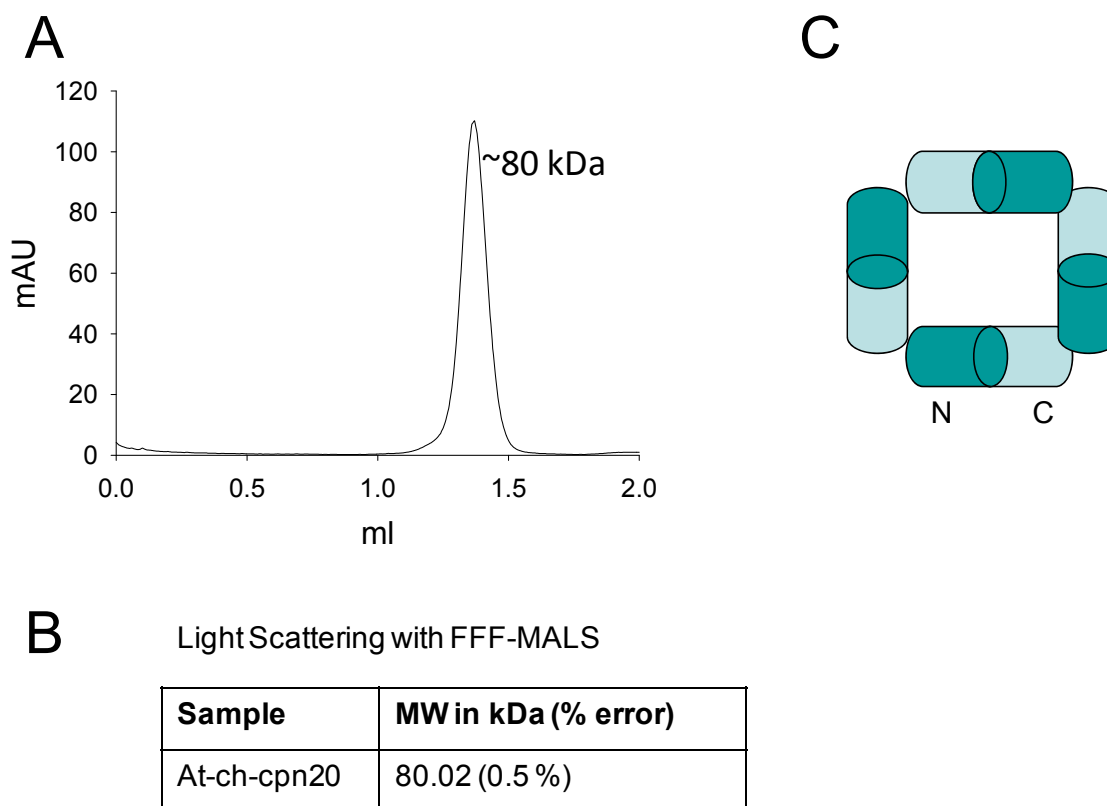
**Figure 22. Averages of end-on and side-on views of electron microscopic images of At-ch-cpn60**

Averages of end-on (A) and side-on (B) views of At-ch-cpn60. The dimensions of the complexes are: symmetrized (sym.) diameter 15.27 nm (A) and 16.5 x 14.8 nm (B) respectively.

### 5.1.2 Electron microscopic analysis of *A. thaliana* chloroplast cpn60/20

The higher plant chloroplast co-chaperone cpn20 (ch-cpn20) comprises two GroES-like domains, fused head-to-tail through a short intervening linker (Figure 23C). It has been reported to form tetramers in solution (Hirohashi et al., 1999; Koumoto et al., 1999). The size exclusion chromatography of purified At-ch-cpn20 and the light scattering experiments confirmed its tetrameric nature (Figure 23A and B).

In order to determine how ch-cpn20 tetramer interacts with the heptameric ring of ch-cpn60, the ch-cpn60/20 complexes were analyzed by electron microscopy (negative staining and cryo-electron microscopy).



**Figure 23. Gel filtration profile and model diagram to show At-ch-cpn20 domain orientation**

(A) The Superdex 200 gel elution profile for the recombinantly purified At-ch-cpn20.

(B) Molecular weight of At-ch-cpn20 as determined by light scattering with precedent field flow fractionation or gel filtration (TSK Super 3000 SW) (Light scattering experiment was performed by Dr. Manajit Hayer-Hartl).

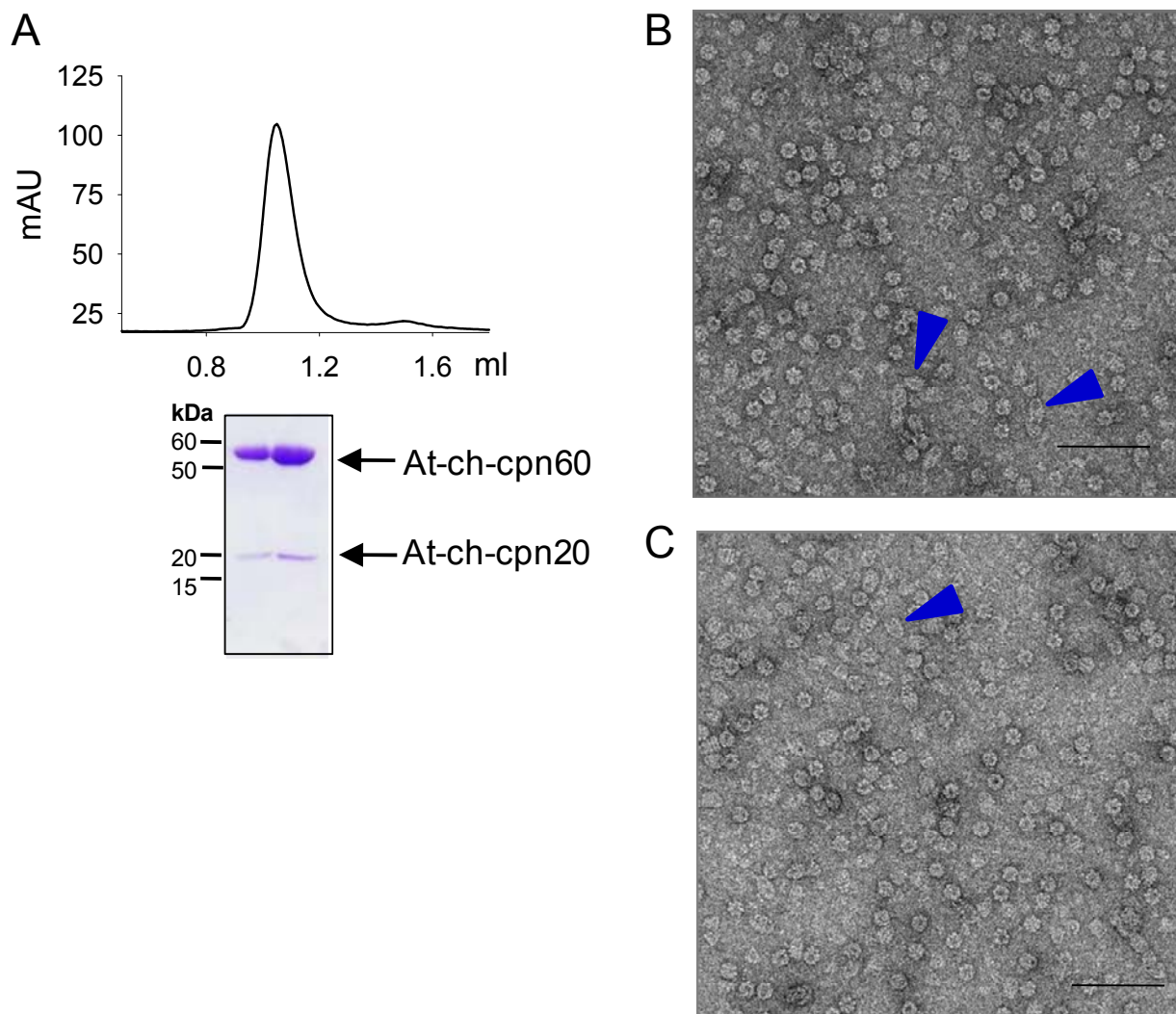
(C) The probable domain orientation of the ch-cpn20 tetramer with the N- and C- terminus labeled.

## RESULTS

---

Recombinantly purified At-ch-cpn60 and At-ch-cpn20 (Figure 19, lanes 5 and 6) were used for the preparation of the chaperone/co-chaperone complex. At-ch-cpn60 and At-ch-cpn20 were incubated in the presence of nucleotide, Mg-ADP or Mg-ATP and the formation of the ch-cpn60/20 complex was analyzed by gel filtration chromatography (Figure 24A). Corresponding to the size of the cpn60/20 complex (~880 kDa), a single peak was observed in the chromatogram. SDS-PAGE analysis of the peak fraction revealed the presence of both ch-cpn60 and ch-cpn20, confirming the formation of cpn60/20 complex. The complexes prepared in the presence of either of the nucleotides were subjected to negative staining using uranyl acetate. Since the complexes prepared in the presence of Mg-ADP showed less number of side-views along with top-views, further experiments were carried out in the presence of Mg-ATP. The images were collected at an instrumental magnification of 28,000 (Figure 24B and C).

From the micrographs, 625 particles with side-on orientation and 1800 particles with end-on orientation were picked manually and subjected to eigen vector-eigen value data analysis. The images obtained during the analysis of side-on views revealed the presence of both cpn60/20 complexes and cpn60 alone (Figure 25A and B). Several bullet-shaped particles were observed, apparently representing asymmetric complexes of the ch-cpn60/20 (Figure 24B and C, blue arrows). These complexes resemble the electron microscopic images of GroEL/ES complexes (Figure 26B), consistent with the binding of cochaperone to the chaperonin asymmetrically (Ishii et al., 1982; Langer et al., 1992b; Saibil et al., 1991). The major class averages obtained during the analysis of side-on views (Figure 26A) revealed a dimension of 19.6 x 14.8 nm, which is close to that of GroEL/ES complexes (19.9 x 13.5 nm) (Figure 26C, (Langer et al., 1992b)).



**Figure 24. Electron microscopic analysis of At-ch-cpn60/20 complex.**

(A) Gel filtration profile for At-ch-cpn60/20 complexes prepared in the presence of Mg-ATP. Below is the SDS-PAGE analysis of the peak fraction showing the presence of both At-ch-cpn60 and At-ch-cpn20.

(B) and (C) Negatively stained images of At-ch-cpn60/20 complex. At-ch-cpn60 and At-ch-cpn20 complexes prepared in the presence of Mg-ATP were stained with 2% uranyl acetate and analyzed with electron microscope at an instrumental magnification of 28,000. The blue triangles show bullet-shaped side-views of At-ch-cpn60/20 complexes. The scale bar represents 100nm.

Averages obtained from the major classes of end-on views revealed that the diameter of the ch-cpn60/20 complex was 15.2 nm, which is very close to the values reported for *E. coli* GroEL/ES (diameter 15.4 nm, (Langer et al., 1992b)). In order to know differences between cpn60/20 and cpn60 end-on views, if there are any, the global

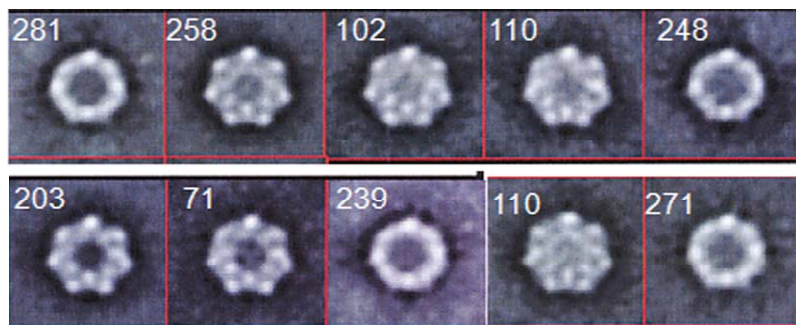


## RESULTS

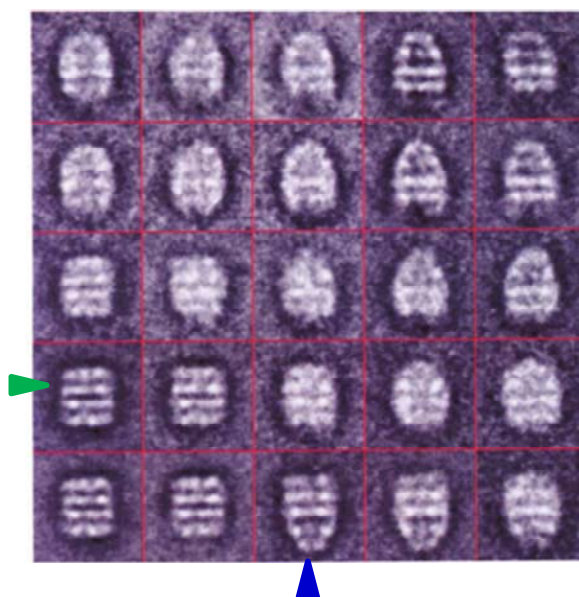
average of cpn60/20 and cpn60 alone were compared (Figure 26D and E). The 7-fold rotational symmetry of cpn60 was preserved in the cpn60/20 complex also. However, a mass apparently representing cpn20 was observed in the center of the cpn60/20 complex (Figure 26D), which was not observed in the global average of cpn60 end-on images (Figure 26E).

However from these images, detailed information about the nature of interaction of the tetrameric ch-cpn20 with the heptameric ring of ch-cpn60 could not be deduced. In order to get further insights into the structure of At-ch-cpn60/20 complexes, the cryo-electron microscopic analysis is being pursued in collaboration with Prof. Wah Chiu at the University of Texas.

A



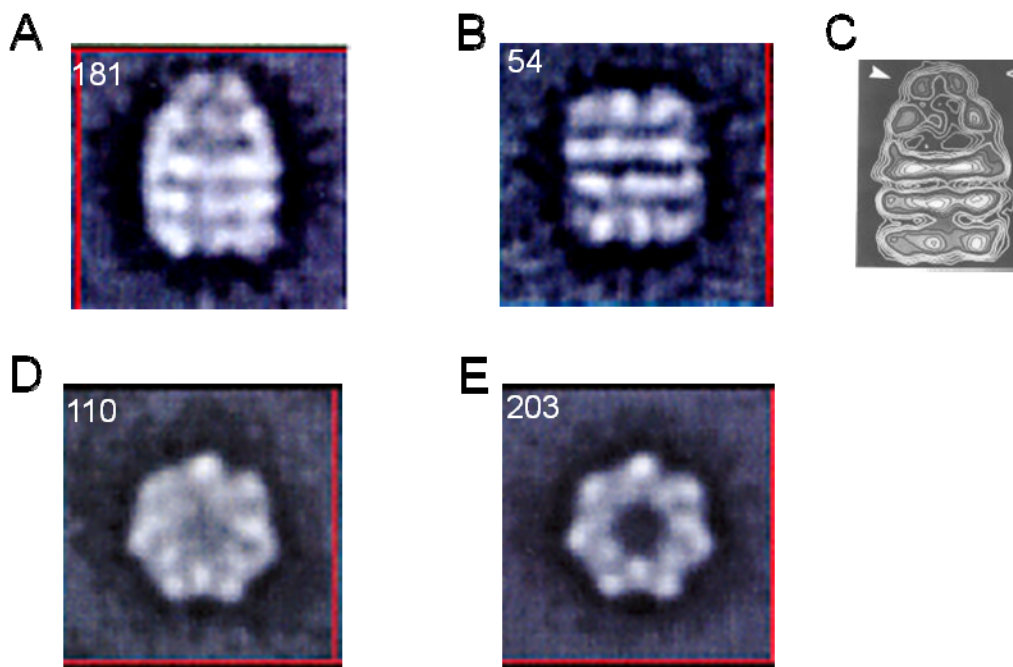
B



**Figure 25. Major classes of Electron Microscopic images for At-ch-cpn60/20 complexes**

(A) Major classes of the top-view of the complexes showing both cpn20 bound and unbound ch-cpn60 particles.

(B) Major classes of the side-view of the At-ch-cpn60/cpn20 complexes (cpn20 bound, blue arrow) and cpn60 (cpn20 unbound, green arrow) particles. Number in the corner depicts the number of particles in that particular class.



**Figure 26. The electron microscopic structure of At-ch-cpn60/20**

(A) and (D) The major averages of side-on (A) and end-on (D) views of the At-ch-cpn60/20 complexes. The dimension of the particle is 19.6 x 14.8 nm (A) and the symmetrized (sym.) diameter is 15.2 nm (D).

(B) and (E) The major averages of side-on (B) and end-on (E) views of the At-ch-cpn60 (cpn20 unbound). The dimension of the particle is 16.5 x 14.8 nm (B) and the symmetrized (sym.) diameter is 15.27 nm (E). Number in the corner depicts the number of particles belonging to the respective class.

(C) Electron microscopic structure of GroEL/ES complex (Langer et al., 1992b)

## 5.2 GroEL induced conformational changes in RbcL

Intrinsic tryptophan fluorescence and I-anilinonaphthalenesulfonate (ANS) binding studies were performed to investigate the conformational changes of Rubisco upon binding and encapsulation by GroEL/ES system.

The changes in the tertiary structure which a protein undergoes upon binding and encapsulation by GroEL can be monitored by observing the intrinsic tryptophan

fluorescence and by ANS binding studies. Most of the intrinsic fluorescence emissions of a folded protein are due to excitation of tryptophan residues, with some emissions due to tyrosine and phenylalanine. Typically, tryptophan has a wavelength of maximum absorption of 280 nm and an emission peak that is solvatochromic, ranging from ~ 300 to 350 nm depending on the polarity of the local environment. ANS is a probe for apolar binding sites whose fluorescence is strongly dependent on the hydrophobicity of the environment. ANS accumulates in the solvated hydrophobic core of early folding intermediates generally termed 'molten globule'. GroEL-stabilized proteins show strong ANS fluorescence (Martin et al., 1991).

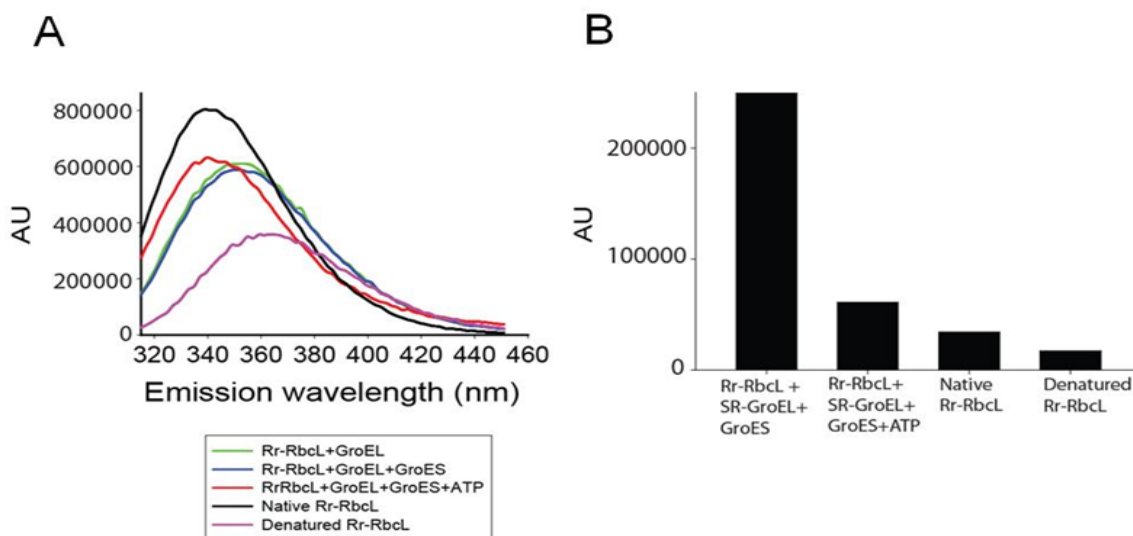
Rubisco large subunit from *Rhodospirillum rubrum* (Rr-RbcL) contains six tryptophan residues and that of *Synechococcus* sp. PCC6301 (Syn6301-RbcL) contains nine. The GroEL and GroES lack any tryptophans (Hemmingsen et al., 1988; Hohn et al., 1979). Single ring version of GroEL (SR-EL) was used for the ANS experiments (Hayer-Hartl et al., 1996; Weissman et al., 1995). SR-EL binds GroES in an ATP-dependent manner, but is unable to dissociate it due to the absence of an allosteric signal from the Gro-EL *trans*-ring. Thus GroES is believed to stably encapsulate protein substrate in the SR-EL/GroES complex without the possibility of active unfolding (Hayer-Hartl et al., 1996).

Native Rr-RbcL showed a maximum emission of tryptophan fluorescence at 335 nm (Figure 27A, black curve), while guanidium denatured Rubisco showed the maximum emission of tryptophan fluorescence at 365 nm (Figure 27A, pink curve) accompanied by a remarkable decrease in the intensity of fluorescence. GroEL-bound Rr-RbcL showed an emission maximum at 348 nm (Figure 27A, green curve & blue curve), i.e. ~57% shift from the denatured to the folded state. High ANS-fluorescence was observed when the unfolded Rr-RbcL was stabilized by SR-GroEL, indicating a massive binding of ANS to the hydrophobic surfaces of the early folding intermediates, in a conformation lacking ordered tertiary structure (molten globule state) (Figure 27B).



## RESULTS

Upon addition of GroEL/ES and Mg-ATP to the unfolded Rr-RbcL, the tryptophan fluorescence emission maximum shifted back to 335 nm, identical to the emission maximum of the native protein (Figure 27A, red curve). This indicates the folding and reconstitution of Rr-RbcL dimer. This was further supported by the decrease in the intensity of ANS fluorescence observed during the refolding reaction (Figure 27B).



**Figure 27. Tryptophan and ANS fluorescence studies of Rr-RbcL upon interaction with GroEL**

A. Tryptophan fluorescence in RbcL refolding: Denatured Rr-RbcL<sub>2</sub> was diluted 100-fold (0.25  $\mu$ M RbcL monomer) into ice-cold assay-buffer containing 0.5  $\mu$ M GroEL. After incubation for 10 min on ice, tryptophan fluorescence (excitation 295 nm, emission scan 315-450 nm) was measured. Then 1  $\mu$ M GroES was added and tryptophan fluorescence was measured again. The reaction was supplemented with 2 mM ATP and refolding kinetics was observed by monitoring the change in tryptophan fluorescence over time. A tryptophan scan at the end of refolding was measured. If only native or denatured substrate had to be analyzed, reactions were modified accordingly. Background fluorescence of chemically identical reactions lacking RbcL was subtracted.

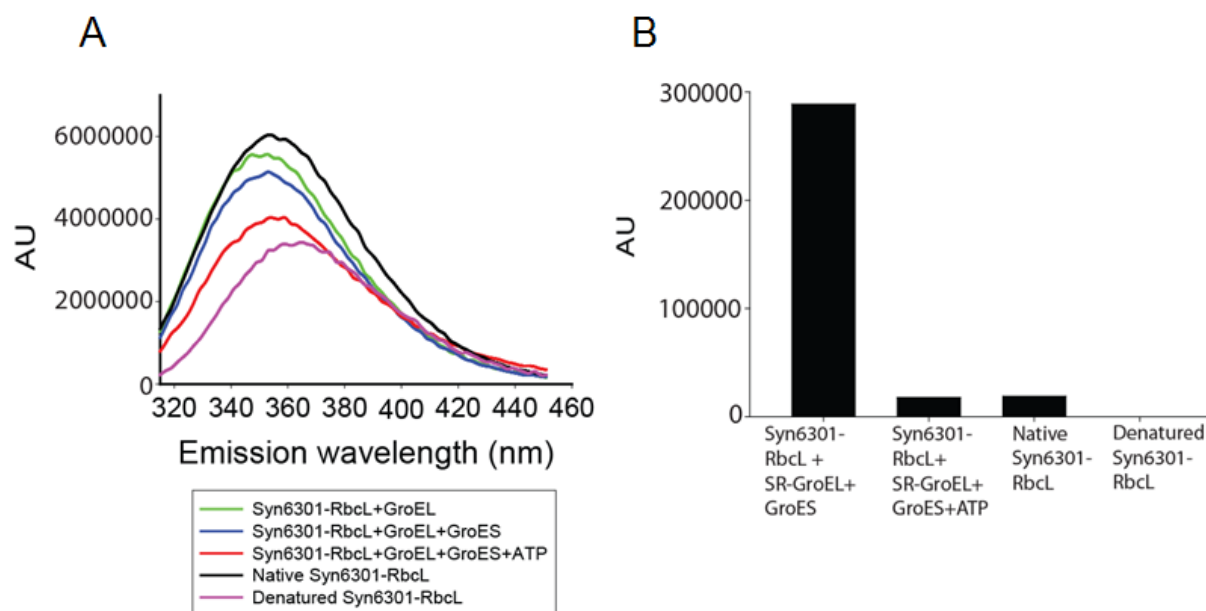
B. ANS-fluorescence in RbcL refolding: Denatured Rr-RbcL<sub>2</sub> was diluted 100-fold (0.25  $\mu$ M RbcL monomer) into ice-cold assay-buffer 1 containing 1  $\mu$ M ANS, 1  $\mu$ M SR-GroEL and 2  $\mu$ M GroES. After incubation for 10 min on ice, ANS fluorescence (excitation 390 nm, emission scan 420-550 nm) was measured. Refolding was started with the addition of 2 mM ATP and the kinetics was observed by monitoring the change in ANS fluorescence over time. At the end of refolding, an ANS fluorescence scan was taken. If only native or denatured substrate had to be analyzed, reactions were modified accordingly. Data were corrected for background fluorescence of chemically identical reactions lacking RbcL and emission at 470 nm was depicted.

Native Syn6301-RbcL<sub>8</sub> showed a fluorescence maximum at 355 nm (Figure 28A, black curve), while that for denatured protein was 365 nm (Figure 28A, pink curve) with a

## RESULTS

decrease in the emission intensity of more than 40%. The emission maximum of GroEL bound Syn6301-RbcL was 355 nm, identical to that of native protein, but fluorescence intensity was slightly lower (Figure 28A, green curve). Similar to Rr-RbcL experiments, high ANS-fluorescence was observed when the unfolded Syn6301-RbcL was stabilized by SR-GroEL (Figure 28B).

Upon addition of GroEL/ES alone or together with Mg-ATP to the unfolded Syn6301-RbcL, the tryptophan fluorescence emission maximum was retained at 355 nm, identical to the emission maximum for native protein with a further decrease in the intensity (Figure 28A, blue curve & red curve). This conformation probably reflects the monomeric folded state of RbcL. The decrease in the ANS-fluorescence again confirms the changes in the tertiary or quaternary structure of the enzyme molecule during the refolding reaction (Figure 28B).



**Figure 28. Tryptophan and ANS fluorescence studies of Syn6301-RbcL upon interaction with GroEL**

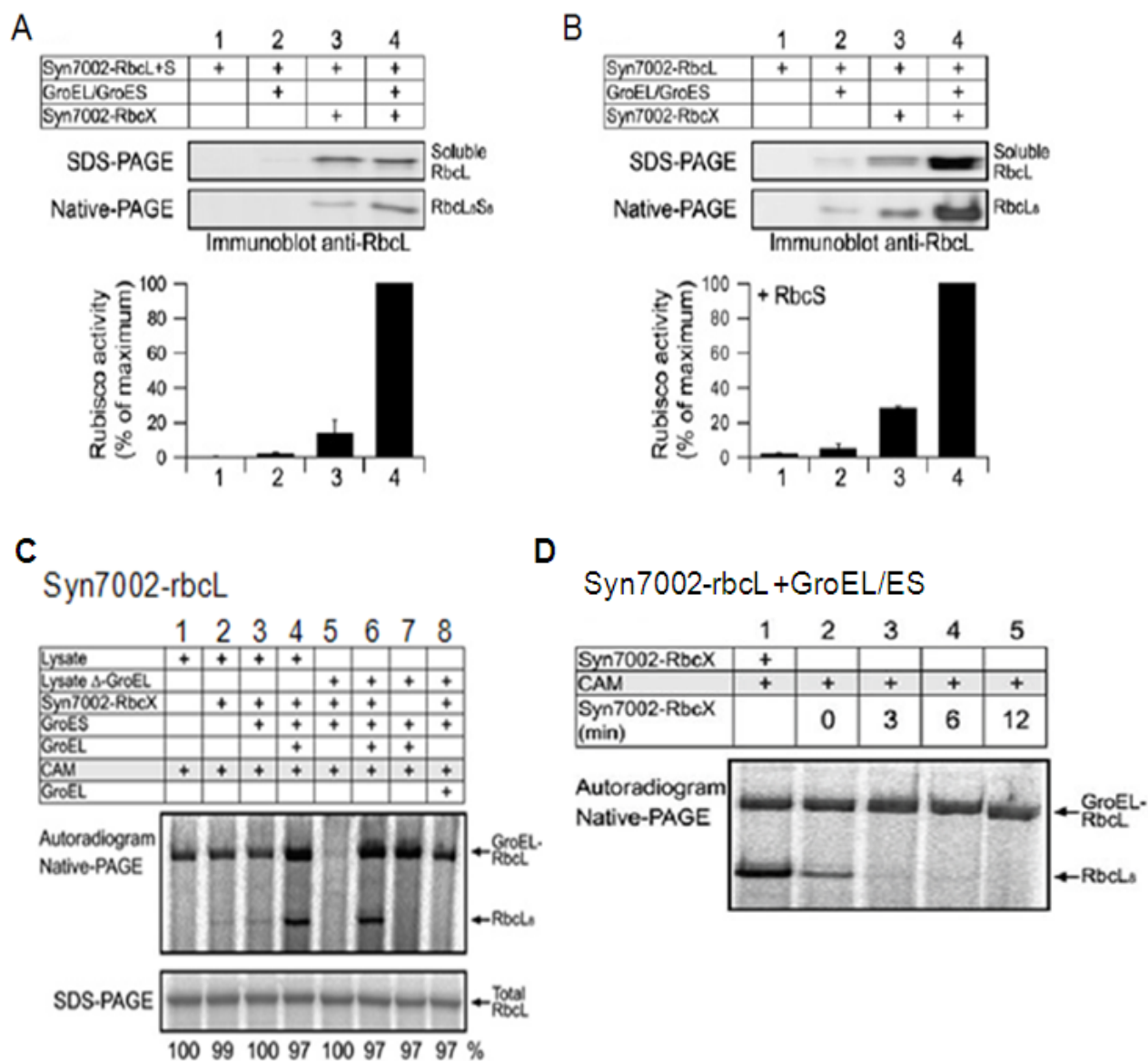
A. Tryptophan fluorescence in RbcL refolding: Denatured Syn6301-RbcL<sub>8</sub> was diluted 100-fold (0.25  $\mu$ M RbcL monomer) into ice-cold assay-buffer containing 0.5  $\mu$ M GroEL. After incubation for 10 min on ice, tryptophan fluorescence (excitation 295 nm, emission scan 315-450 nm) was measured. Then 1  $\mu$ M GroES was added and tryptophan fluorescence was measured again. The reaction was supplemented with 2 mM ATP and refolding kinetics was observed by monitoring the change in tryptophan fluorescence over time. A tryptophan scan at the end of refolding was measured. If only native or denatured substrate had to be analyzed, reactions were modified accordingly. Background fluorescence of chemically identical reactions lacking RbcL was subtracted.

B. ANS-fluorescence in RbcL refolding: Denatured Syn6301-RbcL<sub>8</sub> was diluted 100-fold (0.25  $\mu$ M RbcL monomer) into ice-cold assay-buffer 1 containing 1  $\mu$ M ANS, 1  $\mu$ M SR-GroEL and 2  $\mu$ M GroES. After incubation for 10 min on ice, ANS fluorescence (excitation 390 nm, emission scan 420-550 nm) was measured. Refolding was started with the addition of 2 mM ATP and the kinetics was observed by monitoring the change in ANS fluorescence over time. At the end of refolding, an ANS fluorescence scan was taken. If only native or denatured substrate had to be analyzed, reactions were modified accordingly. Data were corrected for background fluorescence of chemically identical reactions lacking RbcL and emission at 470 nm was depicted.

### 5.3 Requirement of chaperonin system and RbcX for the folding and assembly of Rubisco

As mentioned above, the GroEL/ES system is implicated in the folding of Rubisco (Goloubinoff et al., 1989a; Goloubinoff et al., 1989b; Gutteridge, 1995). *Synechococcus* sp. PCC6301 (also known as *Anacystis nidulans*) Rubisco was expressed as an active holoenzyme in *E. coli* with the help of the GroEL/ES chaperone system (Goloubinoff et al., 1989b). However, the coexpression studies using Rubisco from another cyanobacterium *Synechococcus* sp. PCC7002 (Syn7002) in *E. coli* indicated the GroEL/ES were not alone sufficient for the production of active holoenzyme (Figure 29A and B, Lane 2) (Saschenbrecker et al., 2007). In Syn7002, the product of *rbcX* gene, which is juxtaposed to *rbcL* and *rbcS*, was shown to enhance the production of Rubisco holoenzyme (Larimer and Soper, 1993; Li and Tabita, 1997; Onizuka et al., 2004). Indeed, the *E. coli* expression and the *in vitro* translation studies revealed the requirement of RbcX protein along with the GroEL/ES system. This study also revealed that RbcX interacts with RbcL upon GroEL/ES assisted folding, promoting RbcL to RbcL<sub>8</sub> assembly (Figure 29C). In the absence of RbcX, RbcL forms assembly incompetent aggregates, demonstrated by *in vitro* translation experiments (Figure 29D) (Saschenbrecker et al., 2007).

## RESULTS



**Figure 29. Requirement of chaperonin system and RbcX for production of *Synechococcus* sp. PCC7002 Rubisco**

(A and B) Syn7002-*rbcL* and *rbcS* (A) or Syn7002-*rbcL* (B) was expressed in *E. coli* with or without coexpression of Syn7002-*rbcX* and GroEL/ES, as indicated. RbcL in soluble cell lysates was analyzed by SDS-PAGE and RbcL<sub>8</sub>S<sub>8</sub> or RbcL<sub>8</sub> complexes by Native-PAGE and immunoblotting. Carboxylation activity was measured directly in soluble cell lysates (A) or upon addition of purified RbcS (B). Activities measured upon coexpression with GroEL/ES and RbcX are set to 100%. Error bars indicate standard deviation of three independent experiments (Saschenbrecker et al., 2007).

(C) Translation of Syn7002-*rbcL* was done in *E. coli* lysate *in vitro* in the presence of <sup>35</sup>S-methionine (1.5 hr, 30°C). When indicated, GroEL/ES (0.5 μM/1.0 μM) and Syn7002-RbcX (40 μM) were added to normal lysate (Lanes 1–4) or GroEL-depleted lysate (Δ-GroEL, Lanes 5–8). In lane 8, GroEL was added after

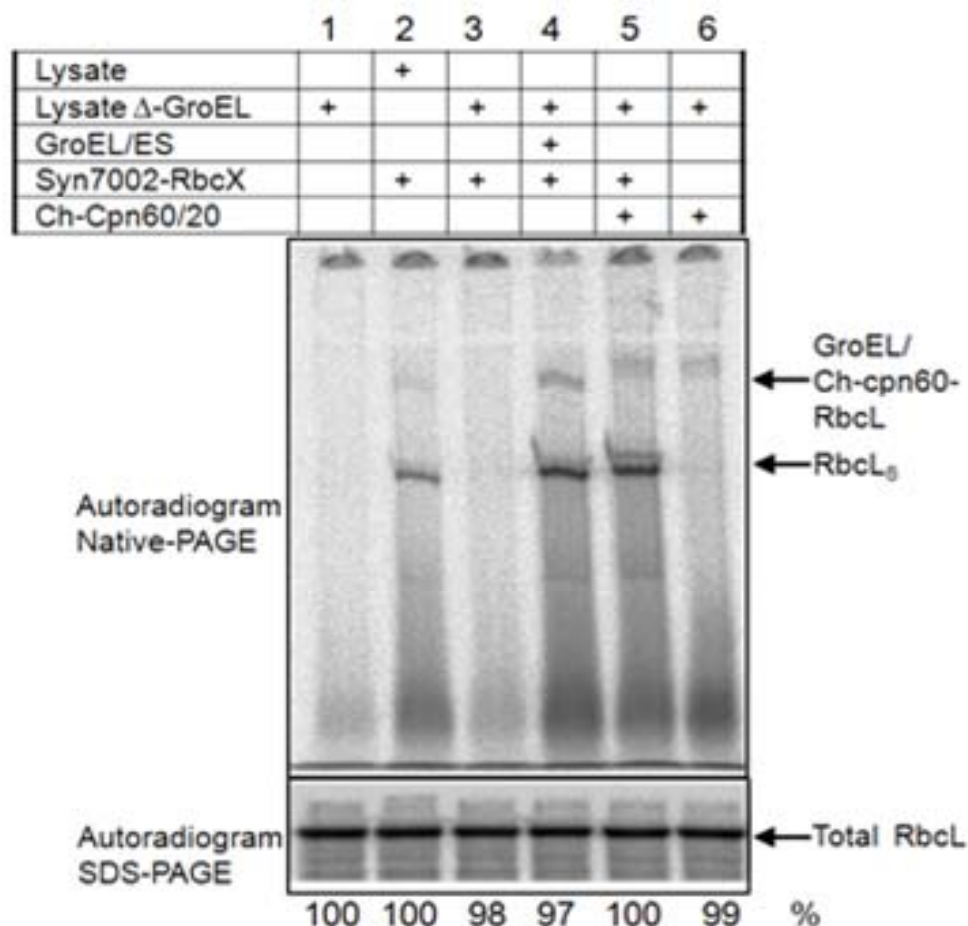
stopping translation with CAM. Assembled RbcL<sub>8</sub> and total RbcL protein were analyzed by Native-PAGE and SDS-PAGE, respectively, followed by autoradiography.

(D) Syn7002-*rbcL* was translated in the presence of GroEL/ES. RbcX was either present during translation (Lane 1) or was added together with CAM (lane 2) or after CAM addition (Lanes 3–5) and assembled RbcL<sub>8</sub> was analyzed by Native-PAGE, followed by autoradiography (Saschenbrecker et al., 2007).

### 5.4 Functional similarity between GroEL/ES and chloroplast-cpn60/20

The overall structural resemblance of *A. thaliana* ch-cpn60 with *E. coli* GroEL and that of At-ch-cpn60/cpn20 complexes with GroEL/ES was evident from the electron microscopic analysis (Figure 22 and 26). Additionally, the reconstitution experiments using Form II Rubisco from *Rhodospirillum rubrum* have shown that the chloroplast-cpn60/cpn20 can functionally replace the bacterial GroEL/ES system (Dickson et al., 2000). In order to analyze if ch-cpn60/20 can functionally replace GroEL/ES, for the folding of cyanobacterial Rubisco, the translation of Syn7002-*rbcL* was carried out in the GroEL-depleted *E. coli* lysate supplemented with RbcX and ch-cpn60/20 (Figure 30, Lane 5). The formation of Syn7002-RbcL<sub>8</sub> complexes demonstrated that both the chaperonin systems share functional similarities consistent with their structural resemblance (Figure 22 and 26).

Syn7002-*rbcL*



**Figure 30. Ch-cpn60/20 can functionally replace GroEL/ES**

Syn7002-*rbcL* was translated in normal or GroEL depleted *E. coli* lysate *in vitro* in the presence of  $^{35}\text{S}$ -Met (90 minutes, 30°C). When indicated, GroEL/ES or chloroplast cpn60/20 (0.5  $\mu\text{M}$ /1.0  $\mu\text{M}$ ) and Syn7002-RbcX (40  $\mu\text{M}$ ) were added to lysate. Assembled RbcL<sub>8</sub> and total RbcL protein were analyzed by Native-PAGE and SDS-PAGE respectively, followed by autoradiography.

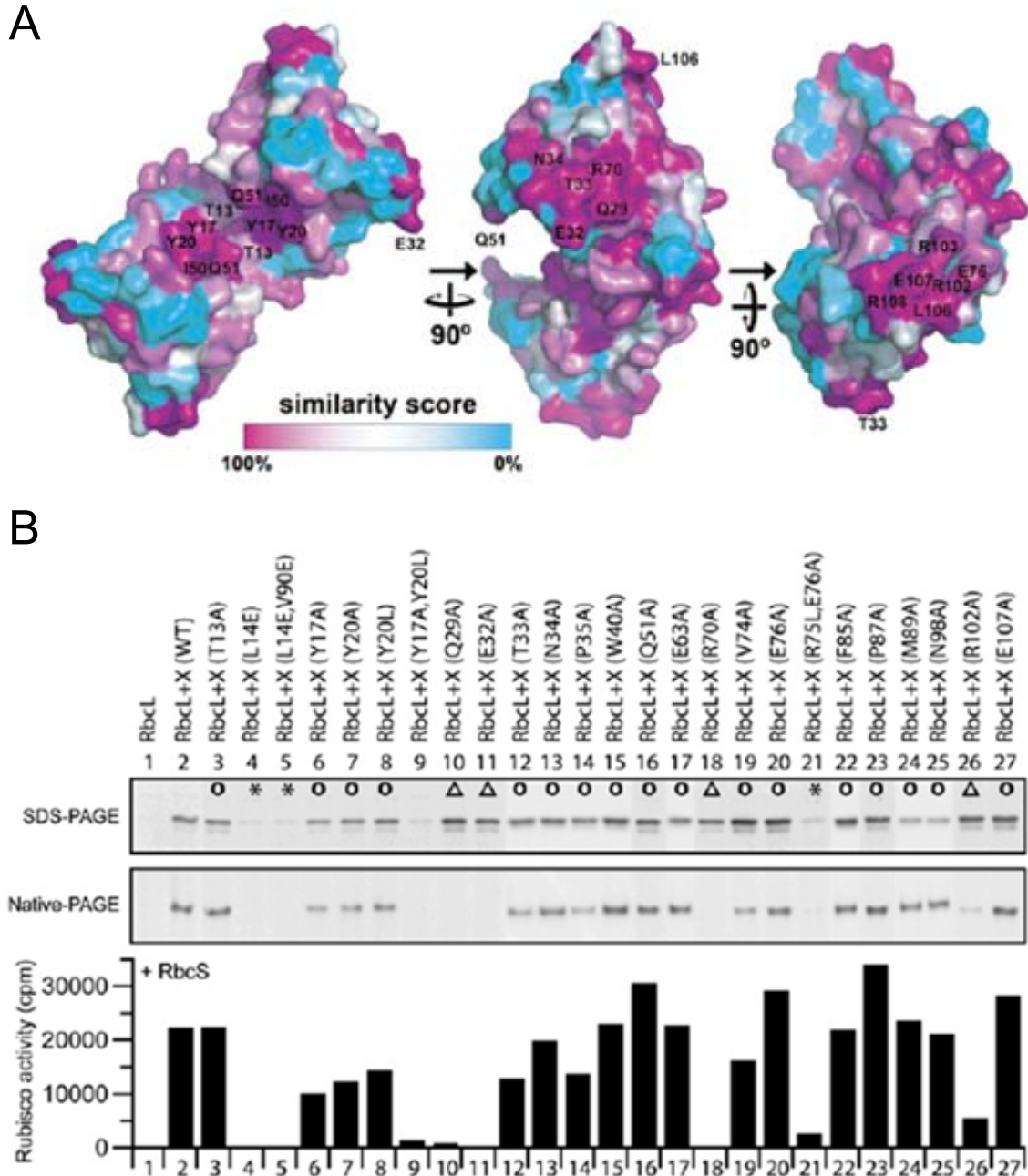
## 5.5 Crystal structure of *Synechococcus* sp. PCC 7002-RbcX mutants

The crystal structure of the wild type Syn7002-RbcX (Saschenbrecker et al., 2007) and the functional analysis of Syn7002-RbcX mutants showed that a hydrophobic area comprising the conserved residues, Y17, Y20, and I50 of each monomer lining the groove in the center of the molecule and the conserved polar residues, Q29, E32 and R70 at the



## RESULTS

corners of the RbcX dimer (Figure 31A) were crucial for the function of RbcX (Figure 31B, lanes 7-9,10-11 & 18) (Saschenbrecker et al., 2007).



**Figure 31. Crucial residues of Syn7002-RbcX**

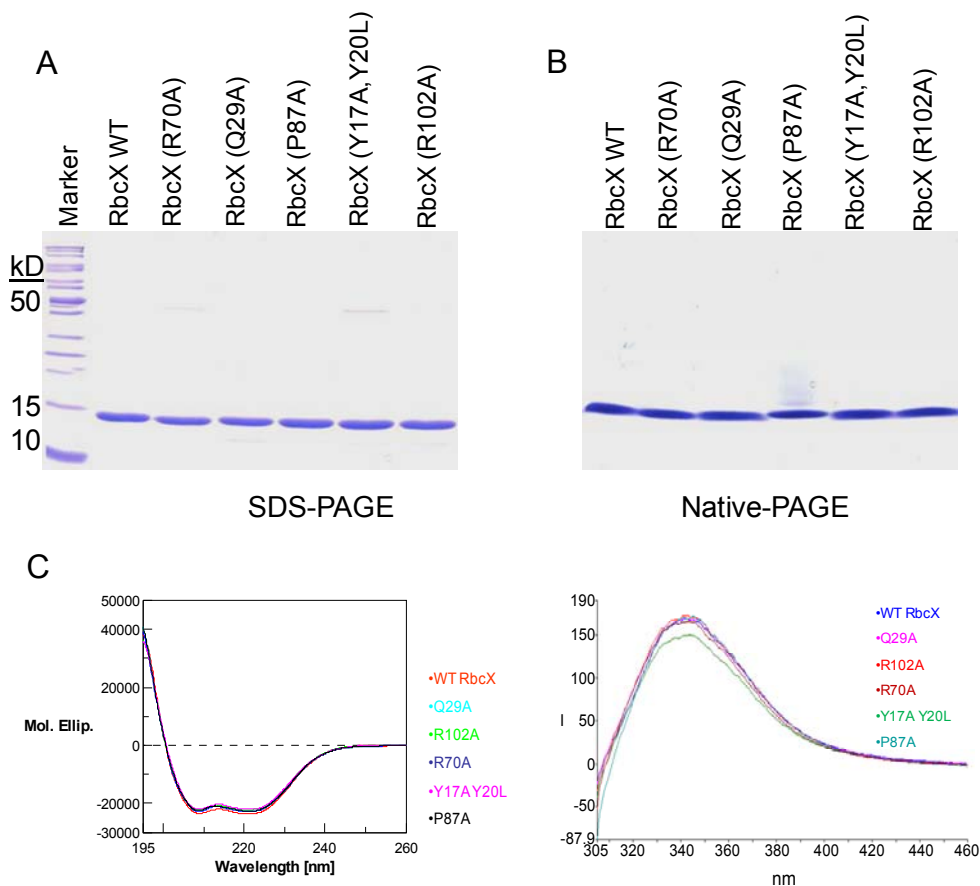
(A) Surface conservation of Syn7002-RbcX. The similarity score from an alignment of 151 sequences of cyanobacterial RbcX in the PFAM database was plotted onto the accessible surface of the RbcX dimer.

## RESULTS

Sequence conservation is indicated by a color gradient, indicating highly conserved residues in magenta and variable regions in cyan. The positions of conserved surface residues are indicated (Saschenbrecker et al., 2007).

(B) SDS-PAGE and Native-PAGE analysis to monitor soluble RbcL and RbcL<sub>8</sub> complex respectively. Solubility of RbcL and formation of RbcL<sub>8</sub> were analyzed upon coexpression of Syn7002-*rbcL* in *E. coli* with wild-type or mutant *rbcX*. (O), RbcX mutants supporting production of soluble and assembled RbcL; (\*), RbcX mutants exhibiting poor solubility; and (Δ), RbcX mutants supporting formation of soluble RbcL but little or no assembly of RbcL<sub>8</sub> and formation of active enzyme upon RbcS addition (Saschenbrecker et al., 2007).

In order to show that the functional defects of RbcX mutants, (Q29A), (R70A) and (Y17A,Y20L), were not a consequence of protein misfolding, we took up crystallization trials for RbcX mutants. We obtained the crystals for these mutants. X-ray diffraction studies and crystal structure analyses were performed in cooperation with Dr. Andreas Bracher. Below is the description of the structural and functional analyses of RbcX mutants and structural comparison with wild type RbcX.



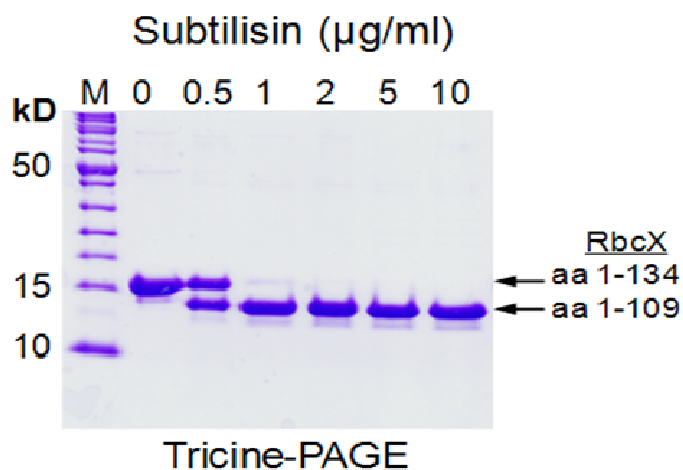


## RESULTS

### Figure 32. Characterization of recombinantly expressed Syn7002-RbcX mutant proteins

(A) and (B) Coomassie stained SDS-PAGE (A) and Native-PAGE (B) of recombinantly purified Syn7002-RbcX wild type (WT) and mutant proteins.

(C) Secondary structure analysis of Syn7002-RbcX mutants. CD analysis (left panel) and melting curve analysis (right panel) of recombinantly purified wild type and mutant RbcX proteins.

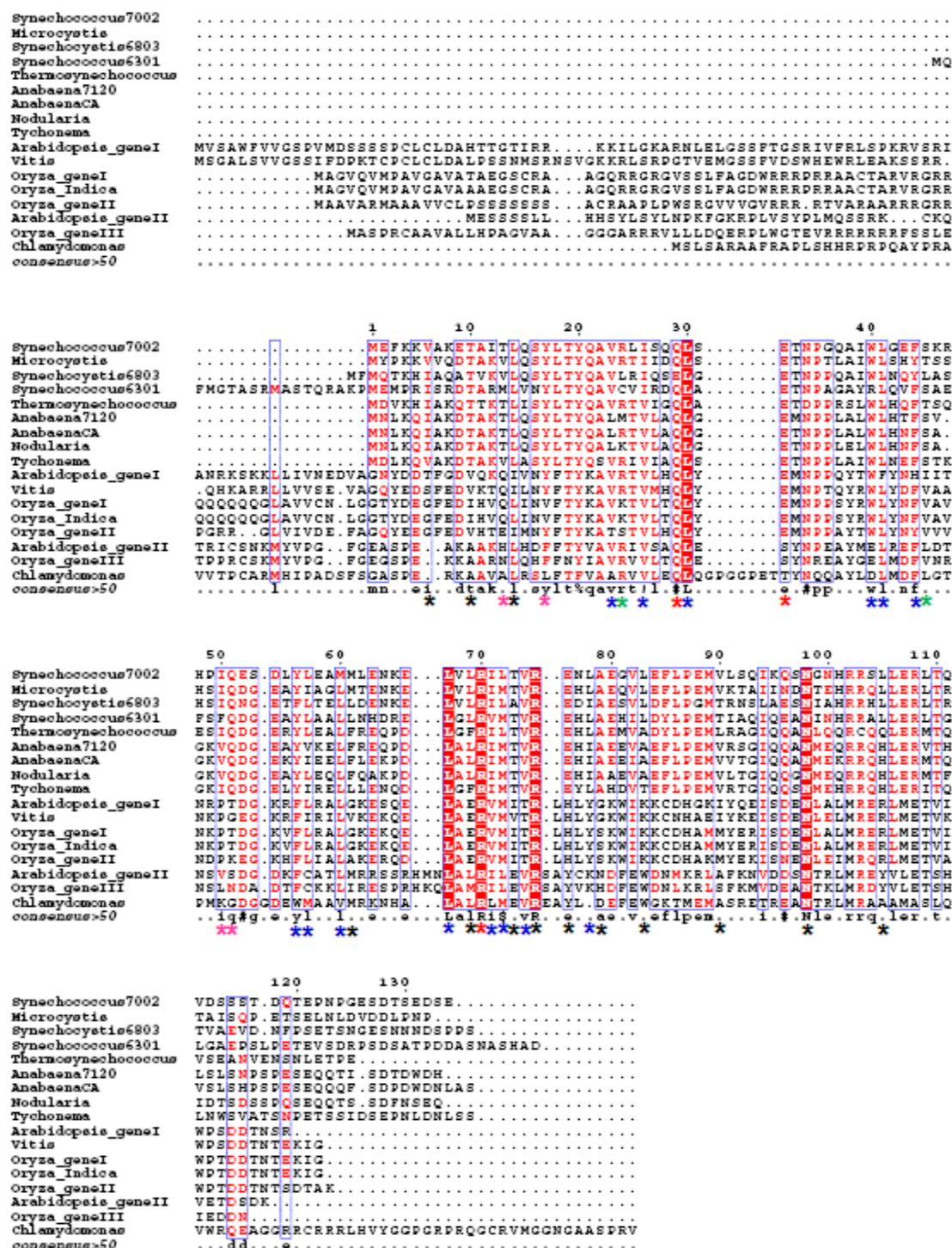


### Figure 33. Limited proteolysis of Syn7002-RbcX

Coomassie stained Tricine-PAGE of subtilisin digested Syn7002-RbcX product.

Wild type or RbcX mutants were expressed recombinantly in *E. coli* and were purified as soluble proteins. On SDS-PAGE, the RbcX wild type and mutant proteins migrated at 15 kDa (Figure 32A). Field flow fractionation in conjunction with multi-angle light scattering (FFF-MALS), revealed that Syn7002-RbcX forms a dimer of 30 kDa (Saschenbrecker et al., 2007). On native-PAGE, the mutant proteins migrated similar to the wild type RbcX dimer indicating that the mutant proteins also formed dimers (Figure 32B). To analyze their secondary structure, the recombinantly purified proteins were subjected to circular dichroism spectroscopy (CD) (Figure 32C). The CD analysis revealed that the secondary elements of these mutants were mostly  $\alpha$ -helical, similar to those of wild type RbcX (Figure 32C). Syn7002-RbcX consists of 134 residues. Limited proteolysis using subtilisin was carried out to identify the potentially unstructured regions in Syn7002-RbcX. It resulted in the deletion of the residues 110-134 at the C-terminus of RbcX (Figure 33). The C-terminal amino acids represent a sequence element that is only poorly conserved among ~150 cyanobacterial RbcX homologs. In contrast, the protease resistant core domain (1-109) is highly conserved (Figure 34).

## RESULTS



**Figure 34. Alignment of amino acid sequence of selected RbcX homologs**

Amino acid sequences of RbcX homologs from *Synechococcus* sp. PCC7002 PR-6 (Q44177), *Microcystis aeruginosa* (Q9R3Q3), *Synechocystis* sp. PCC 6803 (Q55670), *Synechococcus elongatus* PCC6301

(Q5MZ09), *Thermosynechococcus elongatus* (Q8DIS6), *Anabaena* sp.PCC7120 (Q44307), *Anabaena* sp. CA (Q44212), *Nodularia harveyana* BECID29 (Q5K2U3), *Tychonema bourrellyi* (O86944), *A. thaliana*\_gene I (NP\_568382), *Vitis vinifera* (CAO21907), *A. thaliana*\_gene II (NP\_567263), *Oryza sativa* (japonica cultivar-group)\_gene I (Q84M38), *Oryza sativa* (Indica Group) (EAY92274), *Oryza sativa* (japonica cultivar-group)\_gene II (Q7XIH6), *Chlamydomonas reinhardtii* (EDP09788), *Oryza sativa* (japonica cultivar-group)\_gene III (Q7EZV3) were aligned using the Multialign server. Residues indicated in red background are invariable and residues indicated in red are conserved in more than 50% of the sequences; a consensus sequences according to these criteria using the standard symbols of the MultAlin analysis program (Corpet, 1988) is given in the bottom row of the alignment. The secondary structure elements and the residue numbering for RbcX of *Synechococcus* sp. PCC7002 are indicated above the sequences. Residues of the central groove of RbcX containing the phenylalanines and isoleucines in the C-terminal recognition motif of RbcL are indicated by pink and green stars, respectively. Red stars denote residues at the corner surfaces of RbcX implicated by mutagenesis in binding to an uncharacterized surface of RbcL. Black stars denote structural residues involved in dimerization and blue stars indicate residues forming the hydrophobic core of the four-helix bundle.

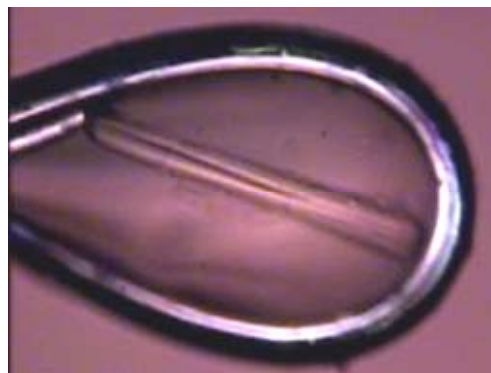
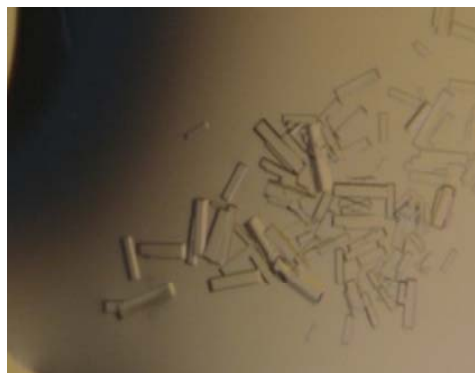
### 5.5.1 Structure of Syn7002-RbcX(Q29A)

Q29 is one of the conserved polar residues at the corners of the RbcX dimer and is necessary for the function of RbcX, as evident from the functional analysis of RbcX (Q29A) mutant (Figure 31) (Saschenbrecker et al., 2007). The corner region of the RbcX dimer is formed by helices 4 of both protomers and by the turn region between the  $\alpha 1$  and  $\alpha 2$  helices of one subunit and may serve to interact with a polar protein surface (Saschenbrecker et al., 2007). This region occurs twice at opposing edges of the dimer due to the 2-fold symmetry.

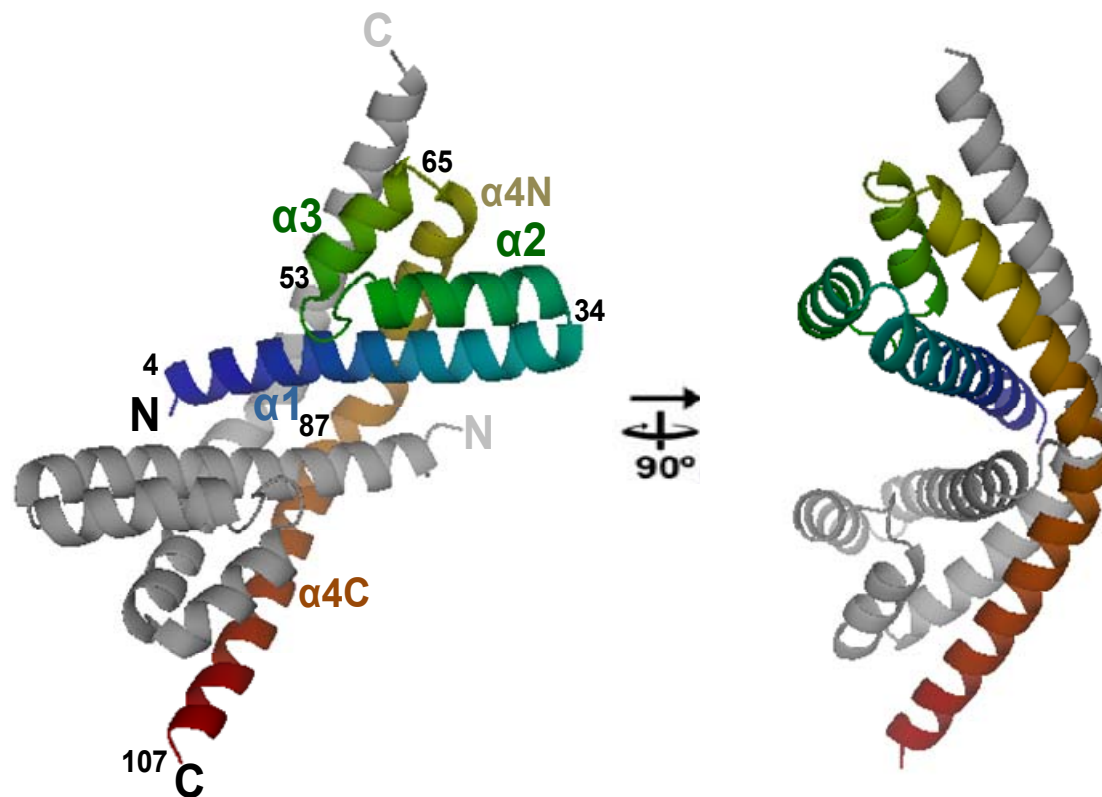
Crystals of Syn7002-RbcX(Q29A) were obtained using the hanging drop vapor diffusion method in the condition, 100mM HEPES-NaOH, pH 7.5, 1.5 M Sodium acetate (Figure 35A). Diffraction data were collected to 3.1 Å resolution at beamline ID29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Diffraction data were integrated with Mosflm (Leslie, 1992) and Scala (Evans, 1997). RbcX(Q29A) crystallized in space group  $P4_12_12$  containing three dimers per asymmetric unit similar to the wild type RbcX (Table 1). While a continuous electron density was observed throughout the core domain of all protein chains, their termini were progressively disordered, presumably due to mobility as indicated by proteolytic susceptibility (Figure

33).  $2F_o - F_c$  electron density for RbcX(Q29A) was calculated after refinement using the wild type RbcX structure as a model. The model was manually rebuilt using Coot (Emsley and Cowtan, 2004) and iteratively refined with Refmac (Murshudov et al., 1997).

A



B



**Figure 35. Crystal Structure of Syn7002-RbcX(Q29A)**

(A) Crystals of Syn7002-RbcX(Q29A) obtained in 100mM HEPES-NaOH, pH 7.5, 1.5 M Sodium acetate. The right panel shows the analyzed crystal mounted in a nylon loop.

## RESULTS

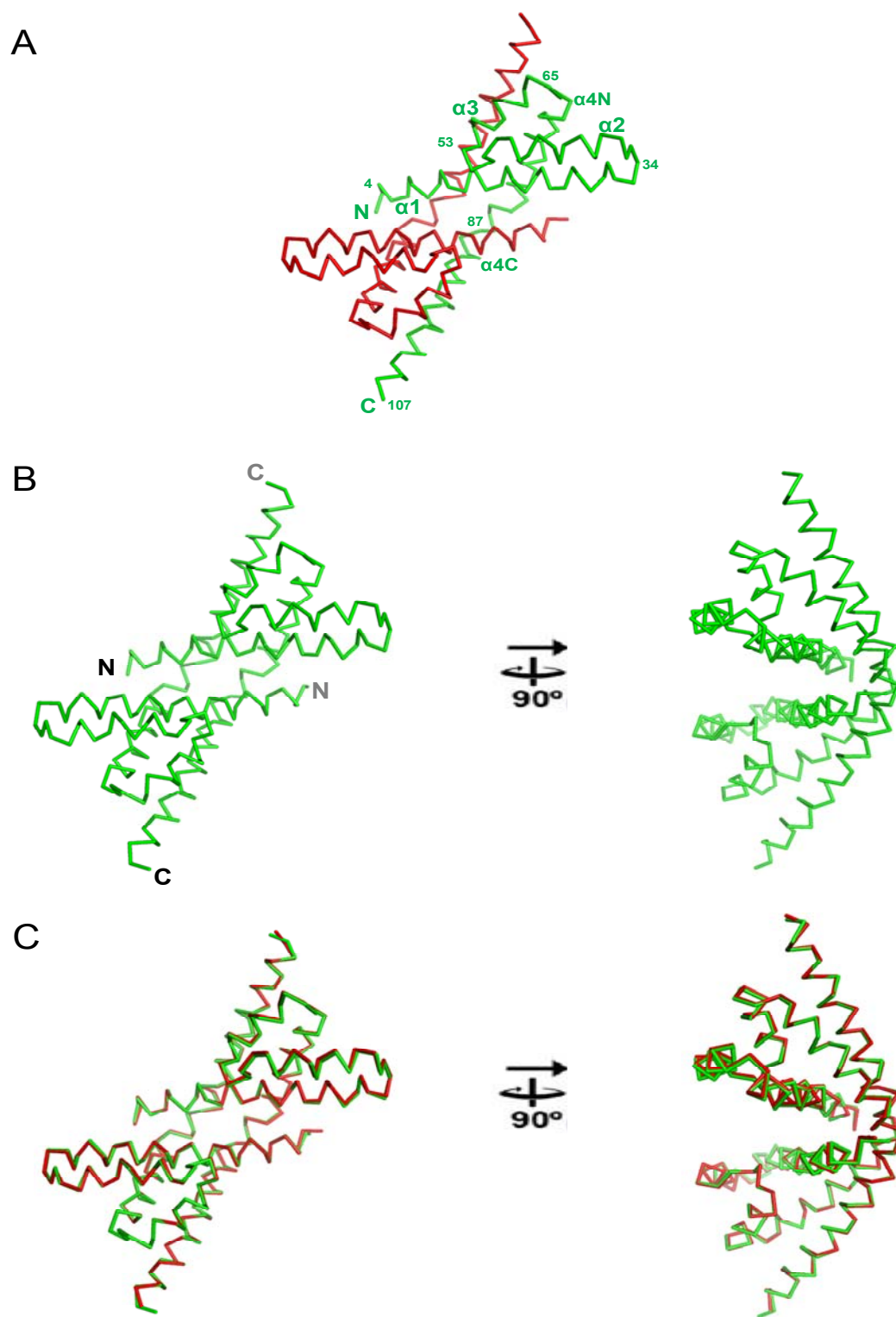
---

(B) Ribbon representation of a RbcX(Q29A) dimer, front and side views. For one of the protomers, the peptide backbone is depicted from N to C terminus using a color gradient from blue to red. Secondary structure elements, selected residue numbers and chain termini are indicated. The other protomer is depicted in gray color and N- and C- terminus are labeled.

The structure of the Syn7002-RbcX(Q29A) core domain shows an helix bundle consisting of four  $\alpha$ -helices ( $\alpha$ 1-4) per monomer (Figure 35B) virtually identical to the wild type RbcX (Figure 36A-C). Helix  $\alpha$ 2 (residues 35-48) turns backward relative to  $\alpha$ 1 (residues 4-33) with a steep interhelical angle of  $151^\circ$  and only one residue (N34) forming the turn. A similar arrangement with an inter-helix angle of  $161^\circ$  is observed at the junction of  $\alpha$ 3 (52-63) and  $\alpha$ 4 (residues 65-107). The central helices  $\alpha$ 2 and  $\alpha$ 3 are comparatively short, comprising only 3 and 2 turns, respectively.  $\alpha$ 2 and  $\alpha$ 3 are connected by a linker of 5 residues. The core of the helical bundle is composed of conserved hydrophobic residues without authentic coiled-coil side-chain packing.  $\alpha$ 4 makes an  $\sim 60^\circ$  kink in the vicinity of residue 84 and forms a 35 Å long extension ( $\alpha$ 4C) pointing away from the helix bundle similar to wild type RbcX (Figure 36A-C).

The RbcX(Q29A) dimer has an overall dimensions of  $\sim 64 \times 33 \times 31$  Å (length X height X width) similar to that of wild type (Figure 36A-C) (Table 1). The long  $\alpha$ 4 helices of the protomers align in an almost antiparallel fashion such that the helical bundles are located at opposite ends (Figure 35B). The  $\alpha$ 1 helices form additional symmetrical contacts and together with the  $\alpha$ 2 helices delineate a narrow diagonal groove in the arc-shaped complex (Figure 35B). Based on the structural comparison with RbcX wild type, it is evident that, the dimer interface is predominantly uncharged and hydrophobic. A polar network around the conserved residue N98, contacting the amide backbone at position L72 and the side chains of R75 and E76 in the opposing protomer, contributes to dimer stability (Saschenbrecker et al., 2007). The  $\alpha$ 2- $\alpha$ 3 helical hairpins make no direct interchain contacts.

## RESULTS



**Figure 36. Comparison of crystal structures of Syn7002-RbcX(Q29A) and Syn7002-RbcX wild type**



## RESULTS

---

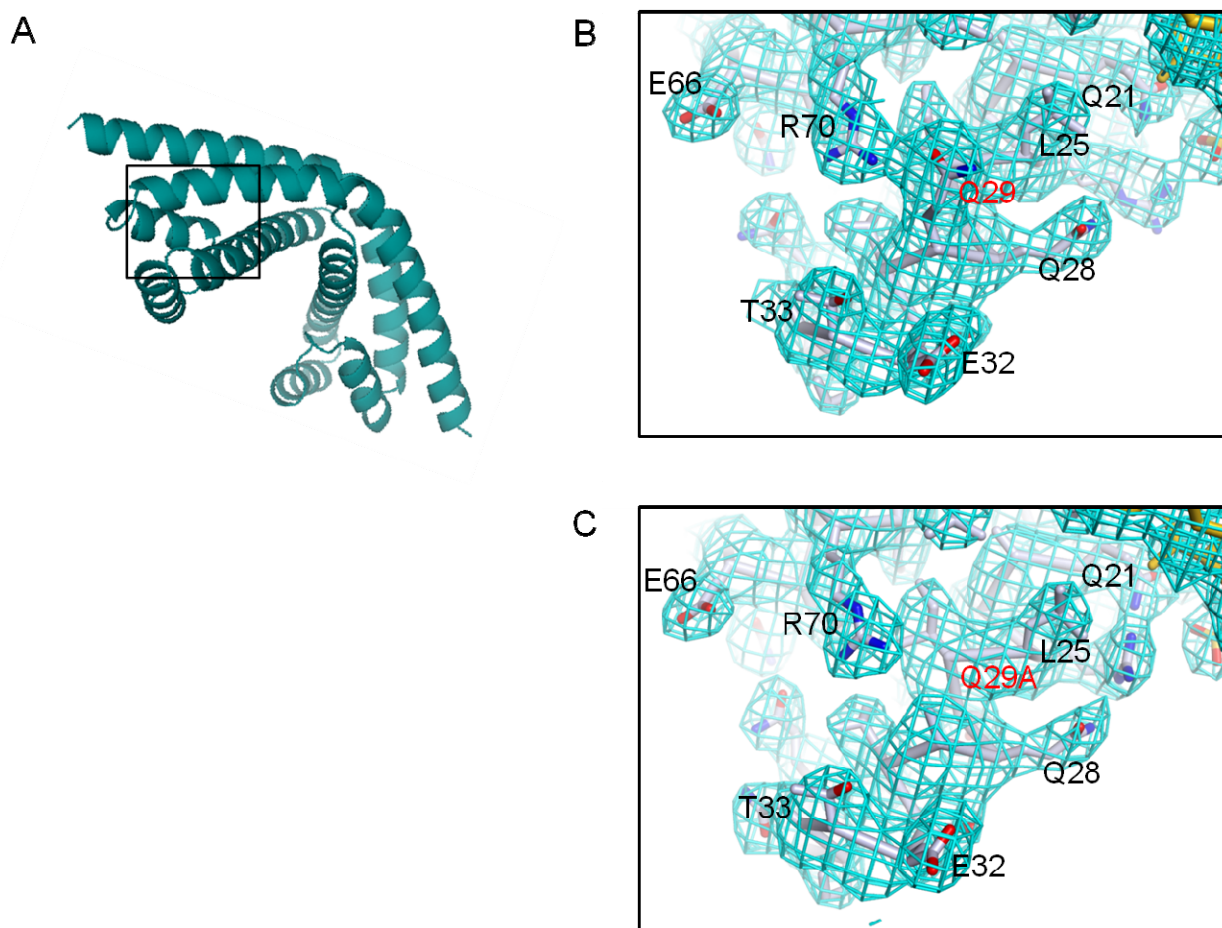
(A) Stick representation of front view of the wild type Syn7002-RbcX dimer. Protomers are shown in green and red. N- and C-terminus of peptide backbone of each monomer is depicted. The selected residue numbers for one of the protomers (green) is shown. Secondary structure elements and chain termini are indicated (Saschenbrecker et al., 2007).

(B) Stick representation of front and side views of the Syn7002-RbcX(Q29A) dimer. Both protomers are shown in green. N- and C-terminus of peptide backbone of each monomer is depicted. The secondary structure elements apply as shown in (A).

(C) Front and side views of the superimposed dimers of Syn7002-RbcX(Q29A) (green) and wild type Syn7002-RbcX (red). The secondary elements apply as shown in (A) and (B).

To further analyze the local alteration of the structure of RbcX upon alanine substitution at position Q29, a comparative analysis of  $2F_o - F_c$  electron density maps of RbcX wild type and RbcX(Q29A) mutant was carried out. In wild type RbcX, the  $\delta$ -amide group of Q29 is in hydrogen bond contact to the guanidinium moiety of R70. Surprisingly, the surface-exposed side chain of the R70 stays in place in RbcX(Q29A) (Figure 37). Apparently the side chain conformation of R70 is mainly governed by hydrophobic packing to the core of the four-helix bundle.

But the effect of alanine substitution at position Q29, disrupting the peripheral polar surface of RbcX, was demonstrated by the coexpression of RbcL and RbcX(Q29A) in *E. coli* (Figure 31B) (Saschenbrecker et al., 2007). Though Q29A mutant allowed accumulation of soluble RbcL, it failed to support the assembly of RbcL to RbcL<sub>8</sub> core complexes (Figure 31B, Lane 10). Further, immunoprecipitation assays confirmed the misassembly and aggregation of RbcL in the presence of RbcX(Q29A) (Saschenbrecker et al., 2007). From these results, it is evident that the peripheral surface of the RbcX dimer is necessary for the proper assembly of RbcL into RbcL<sub>8</sub> core complex.



### Figure. 37 The peripheral polar surface of RbcX

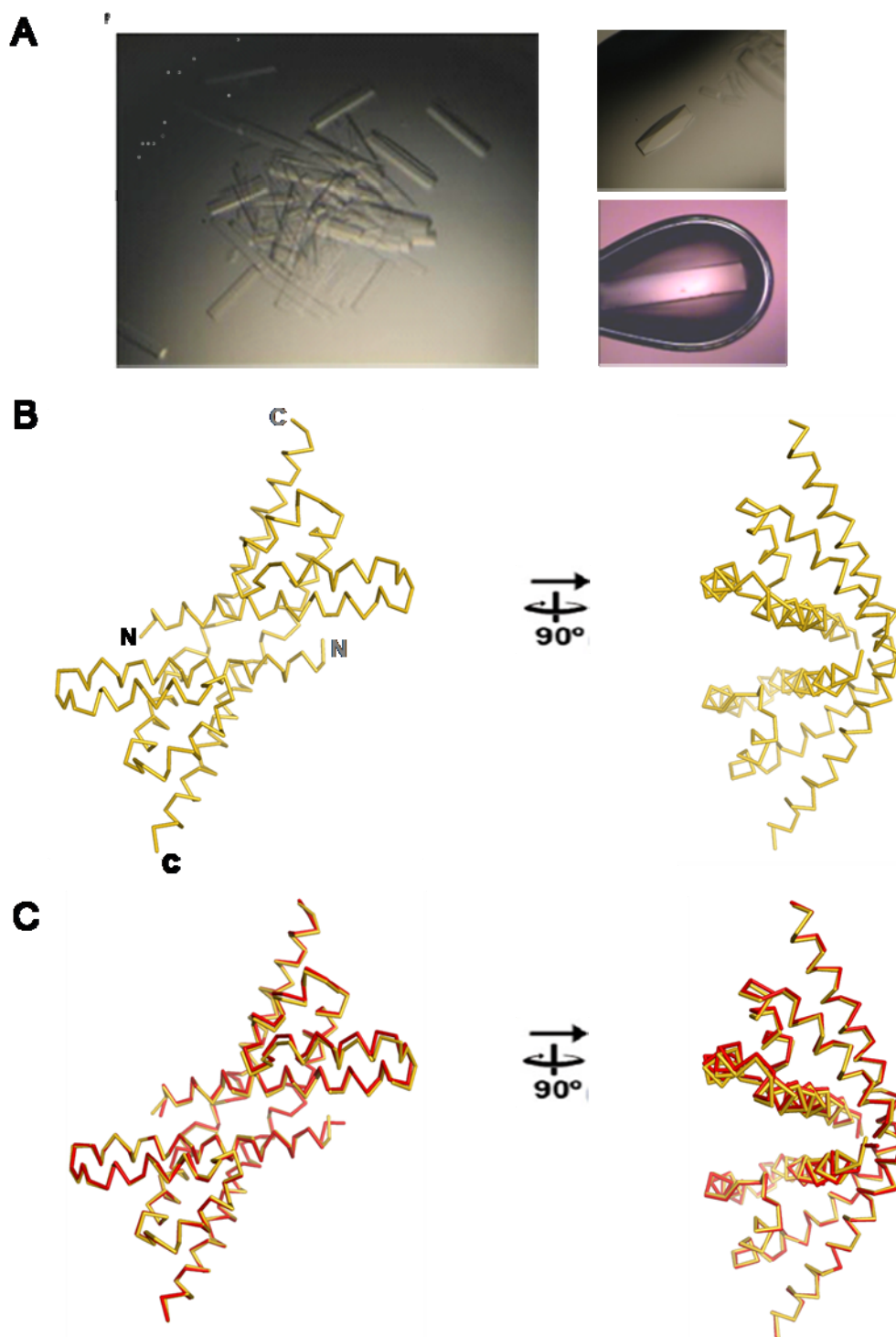
(A) Structure of wild type RbcX dimer. The ribbon representation of the wild type RbcX dimer showing both the protomers in cyan.

(B) and (C) Magnification of the boxed area in (A); Final  $2F_o - F_c$  electron density for the wild type RbcX (B) and RbcX(Q29A) (C) after refinement, contoured at  $1.0\sigma$ . A backbone trace is shown; side chains are represented as sticks and electron density as meshwork in cyan. Carbon, oxygen and nitrogen atoms are indicated in yellow, red and blue respectively.

### 5.5.2 Structure of Syn7002-RbcX (Y17A,Y20L)

Y17 and Y20 are among the conserved hydrophobic residues of each monomer lining the groove in the center of the RbcX dimer. As mentioned above these residues are crucial for the function of RbcX (Saschenbrecker et al., 2007).





**Figure 38. Comparison of crystal structures of Syn7002-RbcX(Y17A,Y20L) and Syn7002-RbcX wild type**

(A) Crystals of Syn7002-RbcX(Y17A,Y20L) obtained in 100mM HEPES-NaOH, pH 7.5, 3M Sodium acetate.

## RESULTS

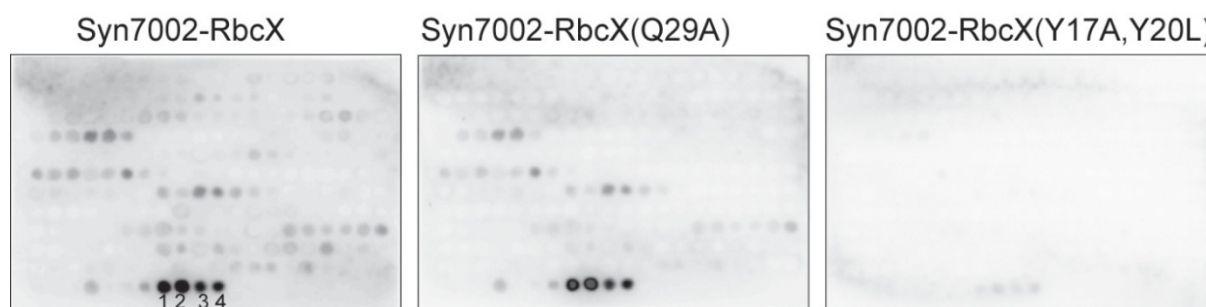
(B) Stick representation of front and side views of the Syn7002-RbcX(Y17A,Y20L) dimer. Both protomers are shown in gold. N- and C-terminus of peptide backbone of each monomer is depicted.

(C) Front and side views of the superimposed dimers of Syn7002-RbcX(Y17A,Y20L) (gold) and wild type Syn7002-RbcX (red).

Crystals of RbcX(Y1A,Y20L) were obtained using the same method as explained for the RbcX(Q29A) mutant. RbcX(Y17A,Y20L) crystallized in the space group  $P4_12_12$  containing three dimers per asymmetric unit (Figure 38A) similar to wild type RbcX and RbcX (Q29A) (Table 1). The RbcX(Y17A,Y20L) structure was solved at a resolution of 3.4 Å. The structure of the Syn7002-RbcX(Y17A,Y20L) core domain was similar to that of RbcX wild type and Q29A (Figure 38B-C).

### 5.5.2.1 RbcL C-terminal peptide recognition by RbcX and the effect of Y17A,Y20L mutations

Peptide binding screens which were performed to identify one or more probable RbcX interacting sequence elements on RbcL revealed that the EIKFEFD close to the C-terminus of RbcL is one of the key binding regions for RbcX (Figure 39 and 42A) (Saschenbrecker et al., 2007). These studies indicated that, while wild type RbcX and RbcX(Q29A) bind to C-terminal peptide of RbcL, RbcX(Y17A,Y20L) fails to bind to C-terminal peptide (Saschenbrecker et al., 2007). To further analyze if there is a local alteration of the structure of RbcX imposed by substituting alanine residues at both the positions Y17 and Y20, a comparative analysis of  $2F_o - F_c$  electron density maps of RbcX wild type and RbcX(Y17A,Y20L) was carried out.



**Figure 39. Interaction of RbcX with the C-terminus of RbcL**

## RESULTS

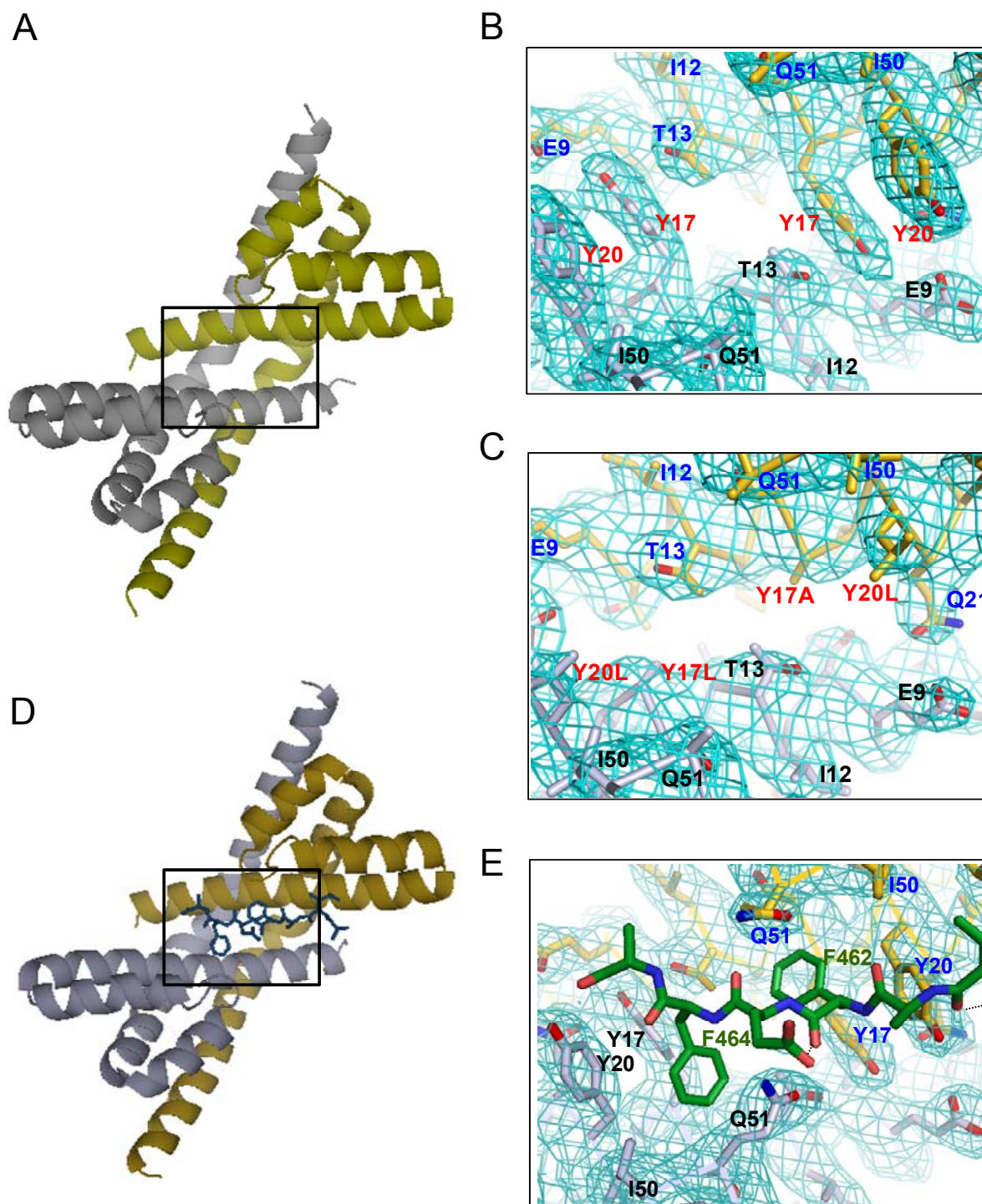
---

Overlapping peptides covering the sequence of Syn7002-RbcL synthesized on a cellulose membrane was probed with the indicated RbcX proteins. Peptide-bound RbcX was visualized by immunodetection with RbcX specific antiserum (Saschenbrecker et al., 2007).

In the wild type RbcX, the central access into the crevice is constricted to a 5.4 Å wide opening by the side chains of the highly conserved Q51 residues that is just wide enough to accommodate a polypeptide chain in an extended conformation, as demonstrated by the crystallographic analysis of EIKFEFD peptide bound to RbcX dimer (Figure 40B and E) (Saschenbrecker et al., 2007). The side chains of F462 and F464 extend into the cavities lined by the side chains of T13, Y17, Y20, I50 and Q51 at the centre of the RbcX cleft (Figure 40E). The polar side chains of the peptide face outwards and are mostly disordered (Figures 40E) (Saschenbrecker et al., 2007). I460 of the peptide appears to contact Y20, R24, and S45 of one RbcX chain (Figure 40E). The conformations of the two Q51 side chains appear to be mainly governed by van der Waals contacts to the phenyl rings of peptide residues F462 and F464. Contact to the phenyl-rings is mediated by hydrophobic packing/exclusion of water against Y17 and Y20 (Figure 40E).

In RbcX(Y17A,Y20L), this crevice is widened extensively which makes RbcX to lose its ability to hold extended polypeptide chain (Figure 40C) and thus the majority of the contacts necessary for RbcL-RbcX interaction are lost in the RbcX(Y17A,Y20L) mutant. Consistent with these observations, in contrast to RbcX(Q29A), RbcX(Y17A,Y20L) failed to support the solubility of RbcL and hence the proper assembly of RbcL to RbcL<sub>8</sub> core complex (Figure 31B, Lane 9).

The mutational analysis of C-terminal RbcL peptide revealed that the phenylalanine residues in the C-terminal peptide of RbcL are crucial for the interaction with RbcX (Saschenbrecker et al., 2007).



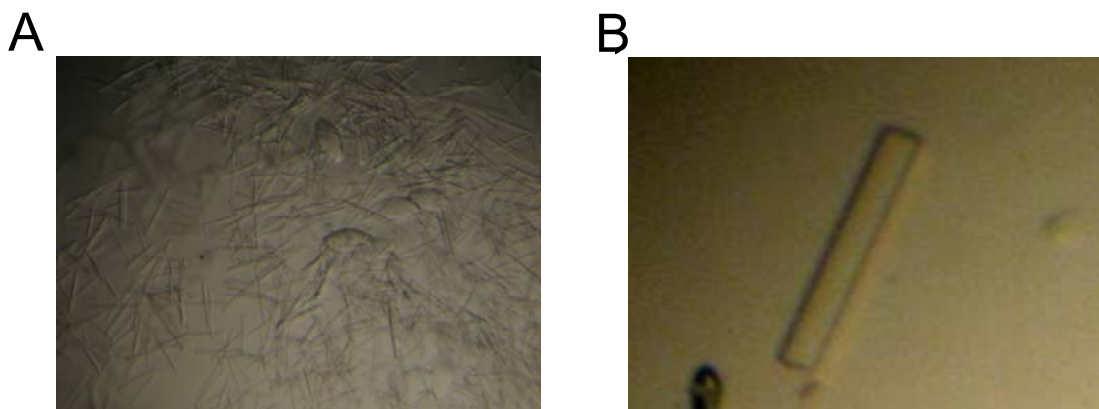
represented as sticks and electron density as meshwork in cyan. Carbon, oxygen and nitrogen atoms are indicated in yellow, red and blue respectively.

(D) Structure of the complex of peptide EIKFEFD bound to RbcX dimer. The peptide shown in stick representation; RbcX is represented as ribbon with protomers coloured in green and cyan, respectively. N and C termini of peptide and each of the RbcX promoters are indicated.

(E) Magnification of boxed area in (A) presenting a view of the refined peptide bound to RbcX. Residues of the RbcX monomers participating in peptide binding are displayed in stick representation and are numbered in green and blue, respectively. The important hydrophobic residues in the bound peptide are also labeled. Dashed lines in black represent hydrogen bonds. A backbone trace is shown; side chains are represented as sticks and electron density as meshwork in cyan (Adapted from Saschenbrecker et al., 2007).

### 5.5.3 Crystallization of Syn7002-RbcX (R70A)

Crystals of Syn7002-RbcX(R70A) were obtained in the condition 100mM HEPES-NaOH, pH 7.5, 2.4 M sodium acetate (Figure 41). The RbcX(R70A) crystals were smaller than the crystals of wild type RbcX and diffracted only to 6.1 Å. The lattice constants were very similar to wild type RbcX crystals (Table 1). An atomic model could not be refined against the RbcX(R70A) dataset because of insufficient experimental parameters limited by the low diffraction power of the crystals. The statistics for the rigid body and TLS refinement of the wild type model against this dataset as given in Table 1 indicate that the crystal packing in the wild type RbcX and RbcX(R70A) was very similar, suggesting that the structure of RbcX is not strongly perturbed by the mutation R70A.



**Figure 41. Crystals of Syn7002-RbcX(R70)**

(A and B) Crystals of Syn7002-RbcX(R70A) obtained in 100mM HEPES-NaOH, pH 7.5, 2.4 M Sodium acetate.



## RESULTS

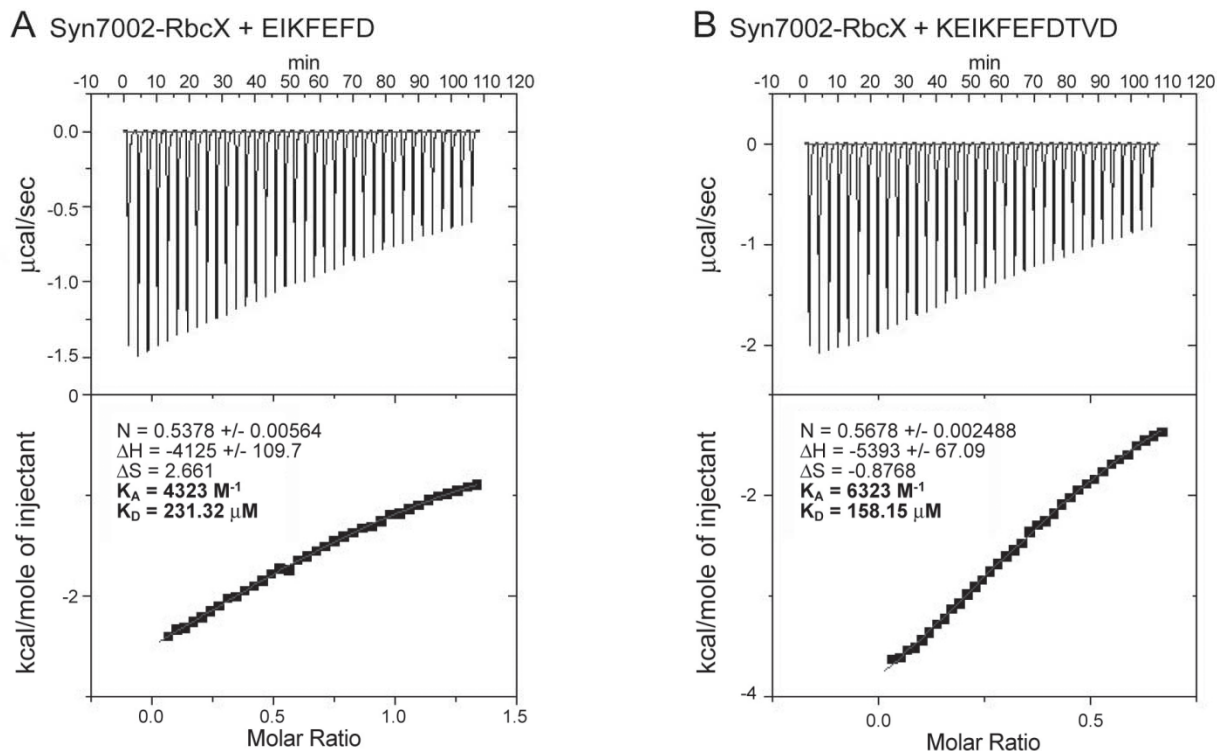
dataset	PtCI4-2	native I	Y17A/Y20L	Q29A	R70A
beamline	ESRF, BM14	SLS, X10SA	ESRF, ID29	ESRF, ID29	ESRF, ID29
Wavelength (Å)	0.8793	0.9789	1.000	1.000	1.000
space group	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2
cell dimensions, a, b, c (Å);	93, 42, 93, 42,	93, 33, 93, 33,	93, 45, 93, 45,	93, 00, 93, 00,	93, 54, 93,
	407.75;	411.99;	411.54;	413.72;	54,
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	413.88;
					90, 90, 90
resolution limits (Å)*	93.25 - 3.15 (3.32 - 3.15)	91.29 - 2.8 (2.95 - 2.8)	93.45 - 3.4 (3.58 - 3.4)	92.85 - 3.1 (3.27 - 3.1)	41.00 - 6.1 (6.43 - 6.1)
Rmerge **	0.081 (0.621)	0.080 (0.641)	0.089 (0.0503)	0.077 (0.485)	0.107 (0.408)
I/sigma	15.9 (3.1)	26.7 (4.6)	9.5 (2.4)	10.6 (2.9)	6.7 (1.8)
multiplicity	7.0 (7.1)	14.5 (14.8)	3.4 (3.6)	3.5 (3.6)	2.0 (1.9)
completeness(%)	100 (100)	100 (100)	99.6 (99.8)	99.9 (99.9)	75.9 (75.9)
RbcX monomers	6	6	6	6	
/a.u.					
solvent content (%)	-	74.7	-	-	
<b>Phasing</b>					
sites	6 (Pt)	-	-	-	
Mean FoM	0.31 (SIRAS)	-	-	-	
<b>Refinement</b>					
resolution range	-	20 - 2.8	20 - 3.4	20 - 3.1	
reflections (test set)	-	43747 (2321)	24596 (1319)	32237 (1726)	
Rwork	-	0.240	0.253	0.231	
Rfree	-	0.262	0.294	0.260	
Number of atoms	-	5105	4727	4976	
r.m.s.d bonds (Å)	-	0.011	0.010	0.010	
r.m.s.d angles (°)	-	1.406	1.319	1.273	
Ramachandran plot (%) ***					
% most favored region	-	93.0	87.8	93.9	
% additionally allowed	-	6.8	11.9	6.1	

Table 1. Crystallographic Data Collection, Phasing and Refinement Statistics

Values in parenthesis are for the highest resolution shell. Data statistics are given according to the definitions used in Scala (Evans, 1997). The Ramachandran plot distributions were determined with Procheck (Laskowski, 1993). An atomic model could not be refined against the RbcX(R70A) dataset because of insufficient experimental parameters limited by the low diffraction power of these crystals. The statistics for the rigid body and TLS refinement of the native I model against this dataset indicated, however that the crystal packing between wild type and R70A mutant RbcX was very similar.

## 5.6 Affinity of RbcX for RbcL peptide

Isothermal titration calorimetry performed to analyze the binding affinities between RbcX and the RbcL C-terminal peptide (Figure 42) showed a relatively low affinity of Syn7002-RbcX for peptide EIKFEFD ( $K_D \sim 230 \mu\text{M}$ ), comprising the core binding region. Slightly higher affinity ( $K_D \sim 160 \mu\text{M}$ ) was measured with the longer peptide KEIKFEFDTVD, which includes the conserved residues Lys458 and Asp468. This indicates that the interaction of Syn7002-RbcX with the RbcL C-terminus is highly dynamic.



**Figure 42. Isothermal Titration Calorimetry of the RbcX-peptide interaction**

Titration of Syn7002-RbcX with Syn7002-RbcL peptide 1. EIKFEFD (A) or peptide 2. KEIKFEFDTVD (B). The stoichiometry of binding is  $\sim 1$  peptide per RbcX dimer (Saschenbrecker et al., 2007)

### 5.7 Role of RbcL C-terminal peptide in the RbcX assisted RbcL<sub>8</sub> assembly

The RbcX interacting RbcL C-terminal peptide is conserved among cyanobacteria (Figure 43A). This peptide sequence is absent in Form II Rubisco (e.g., dimeric Rubisco from *Rhodospirillum rubrum*), (Figure 43C).

To further evaluate the importance of this peptide sequence in the interaction with RbcX, the *rbcL* from either cyanobacteria (*Synechococcus* sp. PCC6301 or *Synechococcus* sp. PCC7002) or *R. rubrum* was translated in the *in vitro* translation system in the presence of Syn7002-RbcX (Figure 43B). Syn7002-RbcX enhanced the production of RbcL<sub>8</sub> core complex for *Synechococcus* sp. PCC6301 (Figure 43B, Lane 5-6), consistent with the presence of RbcL C-terminal peptide and RbcX in this cyanobacterium. Similarly, the translation of *rbcL* from *Synechococcus* sp. PCC7002 resulted in RbcL<sub>8</sub> core complexes only when RbcX was supplemented to the lysate along with GroEL/ES (Figure 43, Lane 7-9). However, Syn7002-RbcX did not enhance the yield of *R. rubrum* dimer (Figure 43B, Lane 2-3), consistent with the absence of RbcL C-terminal peptide and RbcX in this photosynthetic bacterium. These results strongly support the requirement of an assembly chaperone, RbcX for the assembly of complex Form I Rubisco after the GroEL/ES assisted folding of RbcL. In addition, these results reflect that the RbcL C-terminal peptide recognition by RbcX is necessary for the proper assembly of RbcL subunits and this interaction is a universal mechanism in the organisms containing Form I Rubisco such as cyanobacteria and higher plants. The C-terminal sequence of RbcL is located at the surface of the Rubisco holoenzyme and has been implicated in regulating catalysis (Zhu, 1998).



## RESULTS

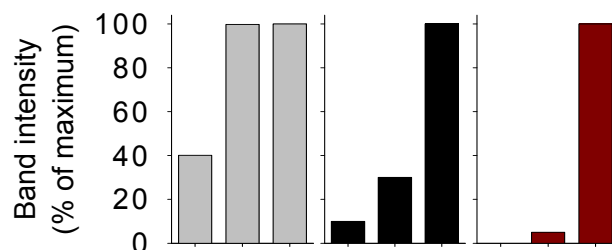
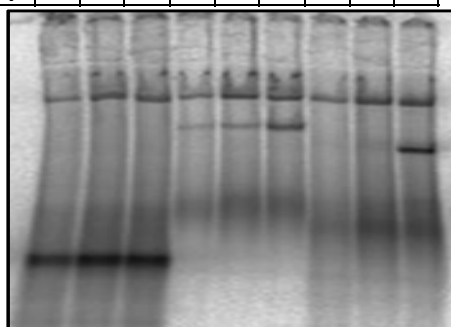
A

Synechococcus7002	420	LEACVQARNEGRSLAREGNDVLRKAGKWSPELAAALDLWKIKFEFDTVDTL
Synechocystis6803	419	LEACVQARNEGRNLAREGNDVIREACRWSPLEAAACELWKIKFEFEAMDTL
Cyanothece7424	421	LEACIQARNEGRSLAREGNDVIREAARWSPLEAAACELWKIKFEFEAMDTL
Microcystis	420	LEACIQARNEGRSLAREGNDVIREACRWSPLEAAACELWKIKFEFEAMDTL
Anabaena7120	425	LEACVQARNEGRNLAREGNDVIREAAKWSPELAVACELWKIKFEFEAMDTV
Thermosynechococcus	424	LEACIQARNEGRDLRMREGGDIIEAARWSPLEAAACELWKIKFEFEAQDTI
Trichodesmium	425	LEACIQARNEGRNLRFREGGDVIREAAKWSPELAVACELWKIKFEFEAMDTL
Lyngbya8106	425	LEACIQARNEGRNLRFREGGDVIREATKWSPELAVACELWKIKFEFEAMDTV
Synechococcus6301	421	LEACVQARNEGRDLRYREGGDILREAGKWSPELAAALDLWKIKFEFEAMDKL
Synechococcus7942	421	LEACVQARNEGRDLRYREGGDILREAGKWSPELAAALDLWKIKFEFEAMDKL
Prochlorococcus9215	416	LEACVKARNAGREIEKESRDILMEAAKHSPLEIALETWKIKFEFOTVDKL
Synechococcus7803	416	LEACVKARNAGRHLEKESRDILMEAGKHSPLEIALETWKIKFEFOTVDKL
Cyanobium7001	416	LEACVKARNAGREIEKESRDILMEAAKHSPLEIALETWKIKFEFOTVDKL
consensus>50		LEAC!qARNeGRnl.rEg.#!iREa.kwSP#LaaAc#lWKEIKFEF#amDtI

B

	1	2	3	4	5	6	7	8	9
Rr-rbcL	+	+	+						
Syn6301-rbcL				+	+	+			
Syn7002-rbcL							+	+	+
GroEL/ES		+	+		+	+		+	+
Syn7002-RbcX			+			+			+

Autoradiogram  
Native-PAGE



C

Rr-RbcL	410	LEQAWCAWRDGVV.LDYAREHKELAFESFPGDADQIYPGMRKALGVEDTRSAFPA
Syn7002-RbcL	415	NRVALEACVQARNEGRSLAREGNDVLRKAGKWSPELAAALDLWKIKFEFDTVDTL
Syn6301-RbcL	417	NRVALEACVQARNEGRDLRYREGGDILREAGKWSPELAAALDLWKIKFEFEAMDKL
At-RbcL	420	NRVALEACVQARNEGRDLAVEGNEIIRKAGKWSPELAAACEVWKIKFTFNFPIDIKLDG
consensus>50		nRvAl#Acv#arnegrdlarEgn#ilReagkwsp#laaalldlWkeikfefdTvdKL

**Figure 43. Importance of RbcX recognition sequence of RbcL**

(A) Alignment using MultAlin (Corpet, 1988) of C-terminal amino acid sequences of RbcL among cyanobacteria. *Synechococcus* sp. PCC 7002 (BAA03076), *Synechococcus elongatus* PCC 6301 (P00880), *Anabaena* sp. PCC 7120 (Nostoc sp. PCC 7120) (P00879), *Synechocystis* sp. 6803 (P54205), *Trichodesmium erythraeum* IMS101 (Q10WH6), *Synechococcus elongatus* PCC 7942 (Q31NB3), *Prochlorococcus marinus* str. MIT 9215 (Q7V6F8), *Synechococcus* sp. WH 7803 (P96486), *Cyanobium* sp. PCC 7001 (CAM91978), *Cyanothece* sp. PCC 7424 (ZP\_02975450), *Lyngbya* sp. PCC 8106 (ZP\_01622123), *Microcystis aeruginosa* NIES-843 (YP\_001659803), *Thermosynechococcus elongatus* BP-

1(NP\_682296). High consensus level ( $\geq 90\%$ ) is depicted in yellow background. Red background shows the conserved residues on C-terminus of RbcL which were detected in peptide screen by using RbcX protein.

(B) Rr-*rbcL*, Syn6301-*rbcL* or Syn7002-*rbcL* was translated in *E. coli* lysate *in vitro* in the presence of  $^{35}\text{S}$ -Met (1.5 hr,  $30^\circ\text{C}$ ). When indicated, GroEL/ES alone or with Syn7002-RbcX were added to the lysate. Assembled RbcL<sub>8</sub> was analyzed by Native-PAGE followed by autoradiography. Below is the quantification of bands on Native-PAGE using Aida-densitometry program. The  $^{35}\text{S}$ -labelled product of translation in the presence of GroEL/ES and RbcX was set to 100.

(C) Alignment using MultAlin (Corpet, 1988) of C-terminal amino acid sequences of RbcL from *Rhodospirillum rubrum* (Q2RRP5), *Synechococcus* sp. PCC 7002 (BAA03076), *Synechococcus elongatus* PCC 6301 (P00880), High consensus level ( $\geq 90\%$ ) is depicted in red background. Yellow background shows the conserved residues on C-terminus of RbcL which were detected in peptide screen by using RbcX protein.

### 5.8 Characterization of *Arabidopsis thaliana* RbcX

Experiments to analyze the function of *Arabidopsis thaliana* RbcX (At-RbcX) were performed in close collaboration with Karnam Vasudeva Rao.

In *Arabidopsis thaliana*, RbcX is encoded by a nuclear gene and RbcL is encoded by a chloroplast gene. At-RbcX shows ~30% sequence identity to cyanobacterial RbcX (Figure 44A). Rubisco large subunits of cyanobacteria and *A. thaliana* share ~80% sequence identity. The EIKFEFD sequence of RbcL C-terminus, important for the RbcX recognition (Figure 39), is conserved in RbcL subunits of higher plants such as *A. thaliana* (Figure 43C), hinting at a possibility of an RbcX mediated Rubisco assembly similar to cyanobacteria. In order to analyze the function of At-RbcX in the Rubisco assembly, the cloning and characterization of At-RbcX was carried out.

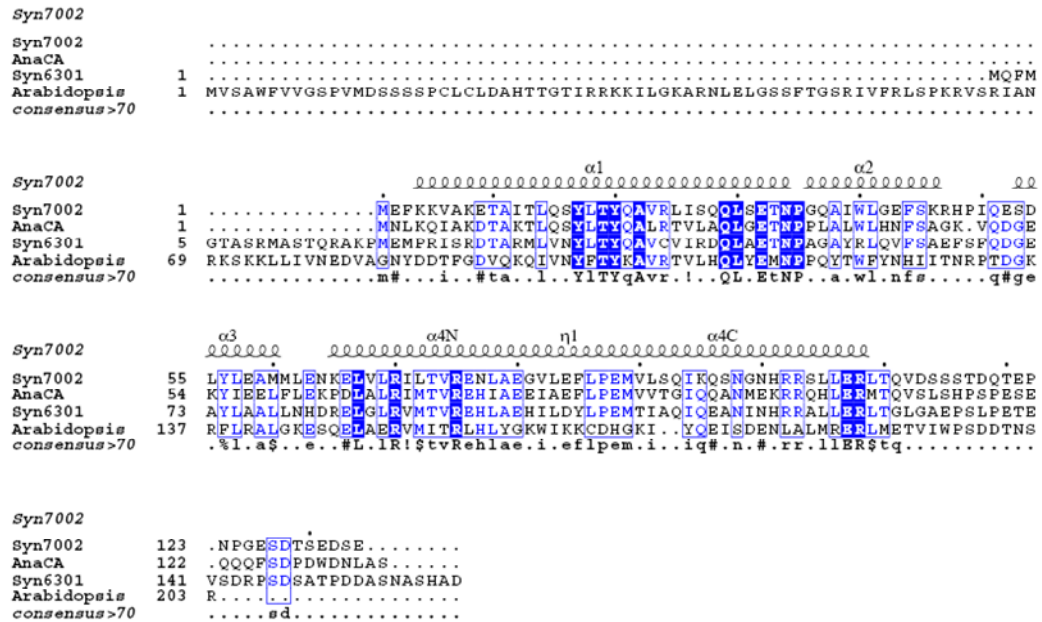
#### 5.8.1 Cloning and purification of At-RbcX

The pHUE vector provides cleavable ubiquitin and His<sub>6</sub> tags and has been extensively used for recombinant expression of broad range of proteins in *E. coli* (Figure 44B) (Catanzariti et al., 2004). At-*rbcX* without the transit peptide was cloned into pHUE vector (Figure 44B). The sequence for the transit peptide was predicted using ChloroP program (Emanuelsson et al., 1999) and confirmed by aligning the amino acid sequences of RbcX proteins from *A. thaliana* and cyanobacteria (Figure 44A). At-RbcX was

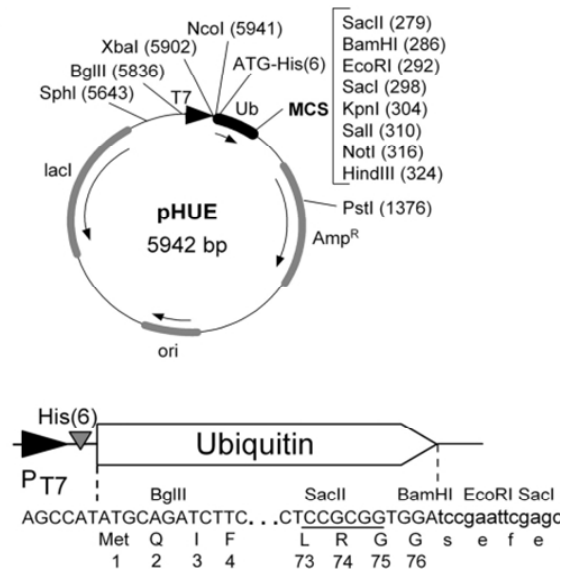
## RESULTS

expressed in *E. coli* and purified by affinity chromatography using Ni-NTA column. At-RbcX runs at an apparent molecular weight of ~15 kDa on SDS-PAGE (Figure 44C). Size exclusion chromatography showed that At-RbcX forms dimers of ~30 kDa similar to Syn7002-RbcX (Figure 45A).

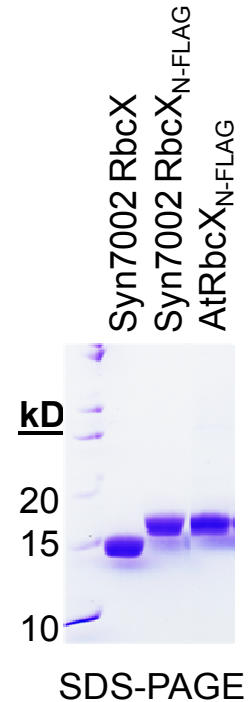
**A**



**B**



**C**



### Figure 44. Cloning and Purification of *A. thaliana* RbcX

(A) Alignment of amino acid sequences of RbcX homologs from *Synechococcus* sp. PCC7002 PR-6 (Q44177), *Synechococcus elongatus* PCC6301 (Q5MZ09), *Anabaena* sp. CA (Q44212) and *A. thaliana* \_gene I (NP\_568382) (with (A) and (B) without transit peptide) using MultAlin analysis program (Corpet, 1988). Residues indicated in blue background are invariable and residues indicated in blue are conserved in more than 50% of the sequences; a consensus sequences according to these criteria using the standard symbols of the MultAlin analysis program is given in the bottom row of the alignment. The secondary structure elements for RbcX of *Synechococcus* sp. PCC7002 are indicated above the sequences.

(B) *At-rbcX* was cloned into *pHUE* vector which provides a cleavable Ubiquitin tag and was expressed in *E.coli* and purified as mentioned in Materials and Methods section (Cloning and purification was done in collaboration with Karnam Vasudeva Rao).

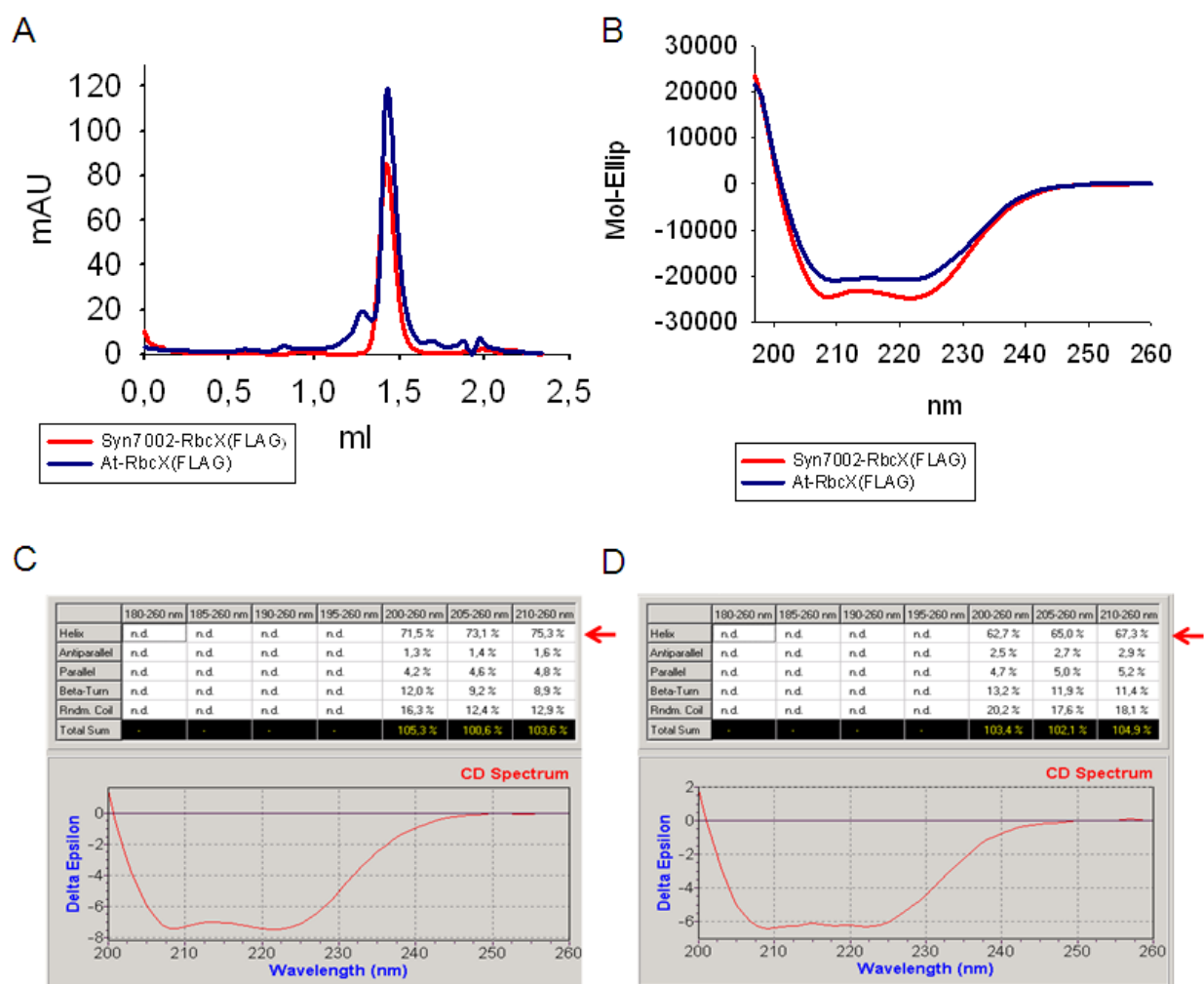
(C) SDS-PAGE followed by Coomassie staining of purified RbcX proteins.

### 5.8.2 Secondary structure and predicted tertiary structure of At-RbcX

Though Syn7002-RbcX and At-RbcX share sequence identity of ~30% only, the functionally important residues mentioned earlier in the study are conserved in At-RbcX also (Figure 44A).

Circular dichroism (CD) studies were performed to analyze the secondary structure of At-RbcX. CD analysis revealed that At-RbcX is mostly alpha helical, similar to RbcX of *Synechococcus* sp. PCC7002 (Figure 45B). The secondary structure predictions, which were performed using the server <http://bioinformatik.biochemtech.uni-halle.de/cdnn>, predicted about 73% and 65% of alpha helical content for Syn7002-RbcX and At-RbcX, respectively (Figure 45C and D).

## RESULTS



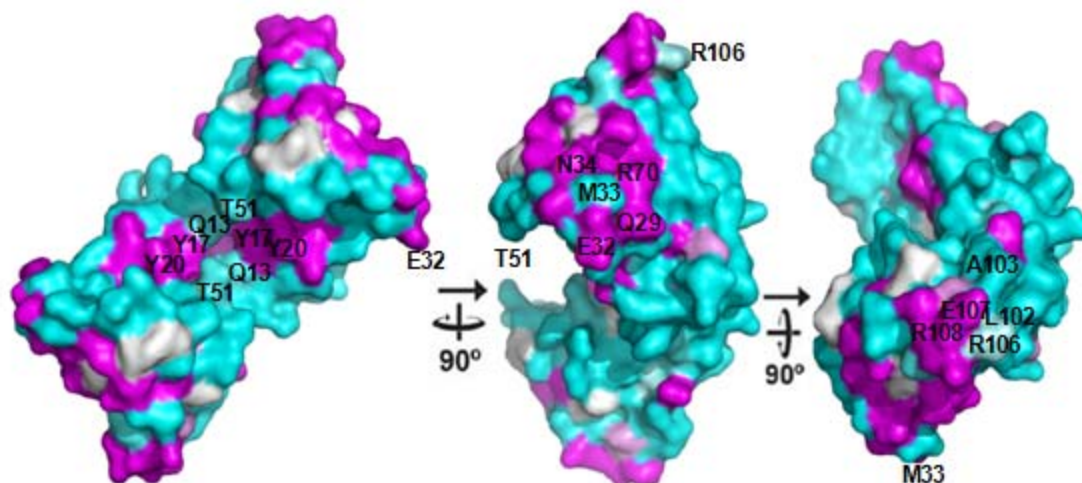
**Figure 45. Size determination and secondary structure analysis of RbcX from *A. thaliana***

(A) Gel filtration analysis of Syn7002-RbcX<sub>FLAG</sub> and At-RbcX<sub>FLAG</sub>.

(B) CD analysis of recombinantly purified At-RbcX<sub>FLAG</sub> and Syn7002-RbcX<sub>FLAG</sub> proteins.

(C) and (D) Secondary structure prediction of Syn7002-RbcX (C) and At-RbcX (D) proteins using cdnn server (<http://bioinformatik.biochemtech.uni-halle.de/cdnn>).

The X-ray crystal structure of Syn7002-RbcX was used as a template to predict the tertiary structure of At-RbcX. The sequence alignment-based superimposition revealed that the polar residues like Q29, E32, R70 located at the corner of the RbcX dimer are conserved in *A. thaliana* RbcX (Figure 46 and 44A). The residues Y17, Y20, and I50 of each monomer, lining the groove in the center of the molecule, are also conserved (Figure 465 & 44A). Most of these residues were shown to be essential for the proper function of Syn7002-RbcX (Figure 31).



**Figure 46. Superimposed Model of *A. thaliana* RbcX**

Surface conservation in At-RbcX. The similarity score from an alignment of amino acid sequences of Syn7002-RbcX and At-RbcX was plotted onto the accessible surface of the RbcX dimer. Sequence conservation is indicated by a color gradient, indicating highly conserved residues in magenta and variable regions in cyan. The positions of conserved surface residues are indicated (These figures were prepared in cooperation with Dr. Andreas Bracher).

### 5.8.3 At-RbcX functions similar to Syn7002-RbcX

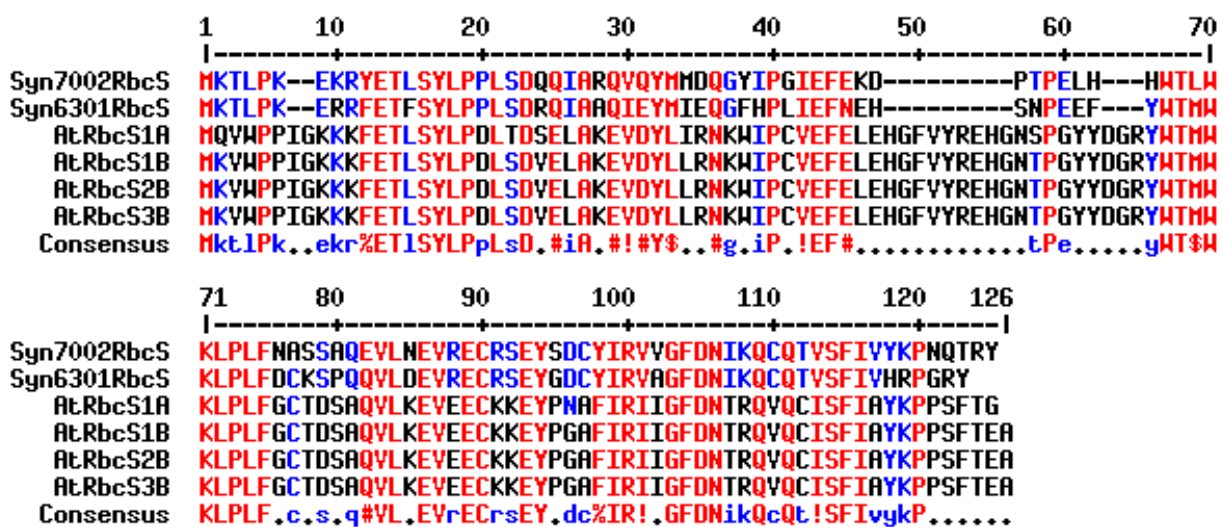
To further analyze the function of At-RbcX, an *in vitro* translation experiment using either Syn7002-*rbcl* or At-*rbcl* and At-RbcX was performed. Earlier to this, RbcS1A of *A. thaliana* used in the translation experiment was cloned and purified as mentioned below.

Four different small subunits exist in *A. thaliana* (RbcS1A, RbcS1B, RbcS2B and RbcS3B), each one encoded by a separate gene in the nuclear genome (Figure 47). The small subunits possess transit peptides, for import into the chloroplasts, which are processed in the chloroplast stroma where assembly of Rubisco takes place (Gatenby and Ellis, 1990). The small subunits of cyanobacteria and higher plants share only 40% sequence identity (Newman and Gutteridge, 1993) (Figure 47). RbcS1A without the transit peptide sequence was cloned into pET vector, expressed in *E. coli* and purified from inclusion bodies.



## RESULTS

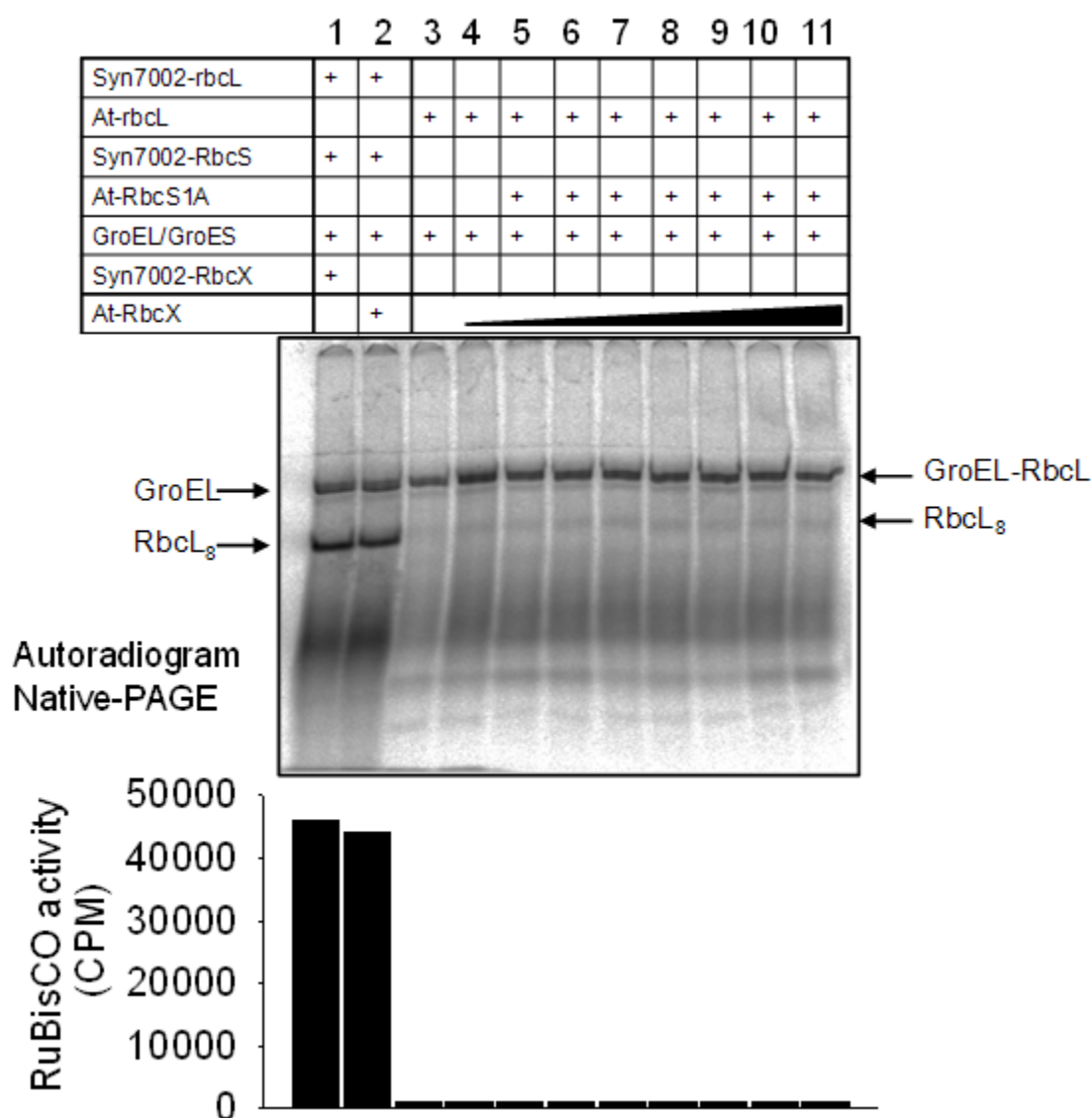
Translation of Syn7002-*rbcL* in the presence of purified Syn7002-RbcS and At-RbcX instead of Syn7002-RbcX, was carried out (Figure 48). Efficient production of active Syn7002-RbcL<sub>8</sub>S<sub>8</sub> complex was observed (Figure 48, Lane 2), similar to the Syn7002-RbcX scenario (Figure 48, Lane 1). In contrast, translating At-*rbcL* in the presence GroEL/ES or At-ch-cpn60/20 and varying amounts of At-RbcX and At-RbcS1A did not yield active Rubisco (Figure 48, Lanes 3-11), suggesting a requirement of an additional factor for the assembly of higher plant Rubisco.



**Figure 47. *A. thaliana* Rubisco small subunits**

Alignment using MultAlin analysis program (Corpet, 1988) of representative small subunits primary structures of Rubisco from *Synechococcus* sp. PCC7002 (Q44178), *Synechococcus* sp. PCC 6301 (P04716), RbcS1A (NP\_176880), RbcS1B (NP\_198657), RbcS2B (NP\_198658) and RbcS3B (NP\_198657) of *A. thaliana* without transit peptide sequences. Residues indicated in red background are invariant and residues indicated in red are conserved in more than 50% of the sequences.

## RESULTS



**Figure 48. Functional Analysis of At-RbcX**

Syn7002-*rbcL* (Lanes 1 and 2) or *At-rbcL* (Lanes 3-11) was translated in *E.coli* lysate *in vitro* supplemented with  $^{35}\text{S}$ -methionine, GroEL/ES (0.5 $\mu\text{M}$ /1.0 $\mu\text{M}$ ), Syn7002-RbcS (14  $\mu\text{M}$ ) (Lanes 1 and 2) or At-RbcS1A (14  $\mu\text{M}$ ) and 20 $\mu\text{M}$  Syn7002-RbcX (Lane 1) or At-RbcX (Lanes 2 and 4-11). Translation was inhibited by the addition of CAM. The soluble products of the translation reactions were analyzed by Native-PAGE, followed by autoradiography. For the assay of carboxylation activity, translation reactions were carried out in the presence of unlabelled methionine and after the CAM addition, products were analyzed for the Rubisco activity (This experiment was performed by Karnam Vasudeva Rao).



## 6 Discussion

Hexadecameric Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) in chloroplasts has been an important paradigm in studies of protein folding and assembly because of its high abundance and functional importance. Rubisco catalyses the initial steps of two competing reactions: photosynthetic carbon fixation ( $\text{CO}_2$  as the substrate) and photorespiration ( $\text{O}_2$  as the substrate) (Andrews and Lorimer, 1987). Photosynthesis results in the synthesis of usable sugars and thus is necessary for plant growth and yield, whereas photorespiration is an energy consuming wasteful process (Andersson and Taylor, 2003).

### 6.1 Structural and functional similarity of chloroplast cpn60/20 to GroEL/ES

Rubisco is one of the stringent substrates of GroEL (Brinker et al., 2001). The bacterial GroEL/ES system and the chloroplast homologue ch-cpn60/20 system are implicated in the efficient folding of Rubisco (Gatenby and Ellis, 1990; Goloubinoff et al., 1989a; Goloubinoff et al., 1989b; van der Vies et al., 1986). The structure of GroEL is well understood (Braig et al., 1994; Langer et al., 1992b, Ranson et al., 2001, Roseman et al., 2001). Whereas GroEL is homooligomeric, ch-cpn60 is heterooligomeric, consisting  $\alpha$  and  $\beta$  subunits (Hemmingsen and Ellis, 1986; Musgrove et al., 1987). However, a detailed structural analysis of ch-cpn60 was yet to be achieved.

The electron microscopic analysis using purified ch-cpn60 from *Arabidopsis thaliana* revealed that the At-ch-cpn60 resembles a 'seven-pointed' star in the top-views (Figure 20A and 20B, red triangles) and a rectangle in the side-on view (Figure 20A and 20B, blue triangles), indicating that the ch-cpn60 has a double toroidal, tetradecameric structure, similar to GroEL (Langer et al., 1992b; Saibil et al., 1991). Averages obtained from the major classes of top views and side views had the diameter and dimension

comparable to those of GroEL, indicating an overall resemblance of ch-cpn60 to GroEL (Figure 22A and B).

GroEL interacts with its co-chaperone GroES in an adenosine nucleotide-dependent fashion (Chandrasekar, 1986; Tilly et al., 1981) and usually forms a bullet-shaped structure in electron micrographs (Ishii et al., 1982; Langer et al., 1992b; Saibil et al., 1991). In contrast to GroEL, the ch-cpn60 has two types of cochaperone, cpn10 and cpn20, the latter consisting of a tandem repeat of cpn10 units (Hill and Hemmingsen, 2001). Unlike the bacterial GroES and the chloroplast cpn10, which are heptameric (Sharkia et al., 2003), the chloroplast cpn20 has been reported to form tetramers (Koumoto et al., 1999).

The electron microscopic analyses of the At-ch-cpn60/20 complexes revealed the presence of several bullet shaped particles representing asymmetric complexes of ch-cpn60/20, similar to GroEL/ES complexes (Figure 24 and 25) (Langer et al., 1992b). In addition, the averages obtained from the major classes of top-views and side-on views of ch-cpn60/20 had the diameter and dimension close to those observed for GroEL/ES complexes (Figure 26) (Langer et al., 1992b). These results indicate that the ch-cpn60/20 complex structurally resembles that of GroEL/ES. Consistently, the translation of cyanobacterial Syn7002-*rbcL* in GroEL-depleted *E. coli* lysate *in vitro*, in the presence of RbcX and At-ch-cpn60/20 resulted in efficient production of RbcL<sub>8</sub> complex, indicating the functional similarity between bacterial and higher plant-chloroplast chaperonins (Figure 30). However from these images, the symmetry mis-match occurring due to the interaction of the tetrameric ch-cpn20 with the heptameric ring of ch-cpn60 could not be explained. To evaluate this aspect, a cryoelectron microscopic analysis of At-ch-cpn60/20 complexes is being pursued.

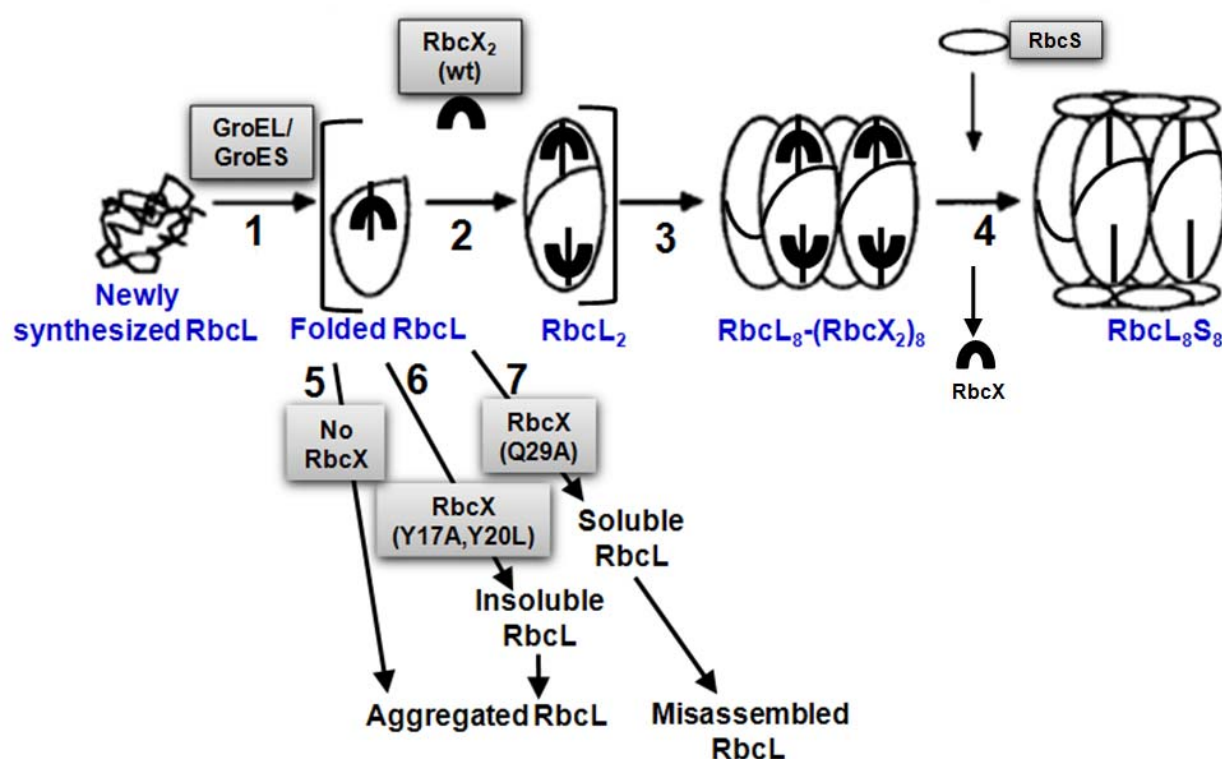
## 6.2 Folding and assembly of Form I Rubisco

The folding and assembly of Form II Rubisco from *Rhodospirillum rubrum* (RbcL<sub>2</sub>) has been reconstituted with the bacterial GroEL/ES system (Goloubinoff et al., 1989a). Similarly, expression of cyanobacterial Form I Rubisco (RbcL<sub>8</sub>S<sub>8</sub>) from *Synechococcus* sp. PCC6301 in *E. coli* was reported to result in the chaperonin-dependent production of enzymatically active holoenzyme (Goloubinoff et al., 1989b). However, *in vivo* and *in vitro* reconstitution of Form I Rubisco from other cyanobacteria and higher plants have not been successful so far (Chaudhari and Roy, 1989; Cloney et al., 1993; van der Vies et al., 1986). Recently, it has been reported that the *rbcX* gene of cyanobacteria enhances the production of enzymatically active cyanobacterial Rubisco upon coexpression with *rbcL* and *rbcS* in *E. coli* (Emlyn-Jones et al., 2006; Li and Tabita, 1997; Onizuka et al., 2004). However, the mechanism of RbcX function was not clearly understood. The structural analysis of RbcX from *Synechococcus* sp. PCC7002, using X-ray crystallography, demonstrated that the RbcX dimer functions as an assembly chaperone for the hexadecameric Rubisco (Saschenbrecker et al., 2007). These studies revealed that the mutations disrupting the conserved peripheral polar surface, (Q29A) or (R70A), or those disrupting the central crevice (Y17A,Y20L) in RbcX cause a defect in the RbcX function (Saschenbrecker et al., 2007).

Consequently, one of the objectives was to analyze in detail the structure of RbcX mutants by X-ray crystallography, to show that the functional defects of RbcX mutants were not a consequence of protein misfolding. RbcX homologues exist in higher plants as well. Another objective was therefore to characterize RbcX of *A. thaliana*.

### 6.2.1 RbcX structure and mechanism

Structural and functional analysis carried out with RbcX wild type (Saschenbrecker et al., 2007) and the present study with RbcX mutants guided us to present the following model for RbcX function (Figure 49).



**Figure 49. Model-RbcX mediated assembly of form I Rubisco**

The newly-synthesized RbcL peptides are folded and released by GroEL/ES system (1). Other chaperones might interact with the RbcL peptides before their interaction with GroEL/ES (1). RbcX (dimer) recognizes the C-terminus of RbcL peptide and binds to it (2). RbcX binds to the early intermediates of RbcL<sub>8</sub> assembly (2) and supports the formation of RbcL<sub>8</sub> core complex (3). Then small subunits assemble individually on top and bottom of the RbcL<sub>8</sub> core complex, thereby displacing RbcX from RbcL<sub>8</sub> to form RbcL<sub>8</sub>S<sub>8</sub> holoenzyme (4). In the absence of RbcX (5) or in the presence of either RbcX(Y17A,Y20L) (6) or RbcX(Q29A) (7), RbcL subunits form aggregates incompetent to form holoenzyme.

RbcX interacts with the RbcL subunits subsequent to their GroEL/ES-mediated folding (Figure 49, step 1). Folded RbcL monomers may spontaneously form dimers or interact with RbcX immediately upon release from GroEL, resulting in stabilization of assembly intermediates competent for efficient progression to RbcL<sub>8</sub> core particles (Figure 49, steps 2 and 3). RbcX recognizes the conserved C-terminal RbcL peptide. The complex between RbcL and RbcX is dynamic, facilitating the eventual displacement of RbcX by RbcS subunits to form the functional holoenzyme (Figure 49, steps 4). In the absence of RbcX, RbcL subunits are prone to aggregation (Figure 49, step 5). In the presence of RbcX with a mutation in the central crevice (RbcX(Y17A,Y20L)), RbcL

subunits are insoluble, forming assembly incompetent aggregates (Figure 49, step 6). Whereas, in the presence of RbcX with a mutation in the peripheral polar surface (RbcX (Q29A)), RbcL subunits misassemble (Figure 49, step 7).

The structural analysis of *Synechococcus* sp. PCC7002 RbcX using X-ray crystallography revealed that the RbcX forms an arc-shaped homodimer (Saschenbrecker et al., 2007). The mutational analysis of RbcX protein identified two critical regions on RbcX, the central hydrophobic crevice and a peripheral polar surface of the molecule. Whereas the central crevice is essential for the production of soluble RbcL subunits (demonstrated by mutations at Y17 and Y20), disruption of the polar surface results in the production of soluble but misassembled RbcL suggesting a role of the RbcX polar surface in proper subunit arrangement of RbcL (demonstrated by mutation at Q29) (Figure 31) (Saschenbrecker et al., 2007).

The crystallographic analysis of Syn7002-RbcX(Q29A) core domain revealed that it consists of a helix bundle with four  $\alpha$ -helices ( $\alpha$ 1-4) per monomer (Figure 35B) similar to wild type RbcX (Figure 36A-C). The crystal structure of Syn7002-RbcX(Q29A) was virtually identical to the structure of wild type Syn7002-RbcX. In wild type RbcX, the  $\delta$ -amide group of Q29 is in hydrogen bond contact to the guanidinium moiety of R70. Surprisingly, the surface-exposed side chain of the R70 stays in place in RbcX (Q29A) (Figure 37). Apparently the side chain conformation of R70 is mainly governed by hydrophobic packing to the core of the four-helix bundle. The functional analysis revealed that RbcX(Q29A) allowed accumulation of soluble RbcL, however, it failed to support the assembly of RbcL to RbcL<sub>8</sub> core complexes (Figure 31B, Lane 10). Furthermore, immunoprecipitation assays confirmed the misassembly and aggregation of RbcL in the presence of RbcX(Q29A) (Saschenbrecker et al., 2007), indicating that the peripheral surface of the RbcX dimer promotes the proper assembly of RbcL into RbcL<sub>8</sub> core complexes.

The structure of the RbcX groove mutant (Y17A,Y20L) revealed that its core domain is similar to that of RbcX wild type and RbcX(Q29A) (Figure 38B-C). In the wild type

RbcX, the central access into the crevice is constricted to a 5.4Å wide opening by the side chains of the highly conserved Q51 residues; that is just wide enough to accommodate a polypeptide chain in an extended conformation (Figure 40B and E) (Saschenbrecker et al., 2007). Whereas, in RbcX(Y17A,Y20L), this groove is widened extensively which makes this mutant to lose its ability to hold extended polypeptide chain (Figure 40C). Most of the contacts necessary for the binding of RbcL C-terminal peptide are lost in the RbcX(Y17A,Y20L). Consistent with these observations, RbcX(Y17A,Y20L) failed to support the solubility of RbcL and hence the proper assembly of RbcL to RbcL<sub>8</sub> core complexes (Figure 31B, Lane 9).

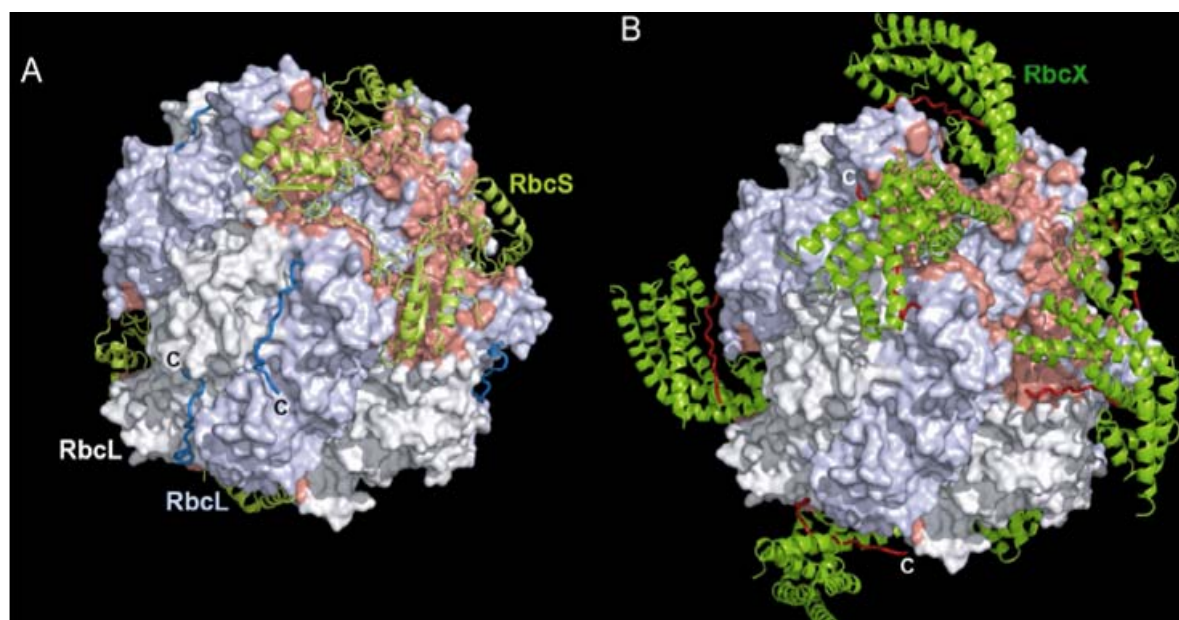
These results revealed that, the point mutations disrupting a peripheral polar surface or the central cleft of RbcX caused a defect in RbcX function and that these defects were not a consequence of protein misfolding.

### **6.2.2 Importance of RbcL C-terminal peptide**

The EIKFEFD motif of the RbcL C-terminus was shown to be a site for RbcX interaction (Figure 39 and 43A) (Saschenbrecker et al., 2007). The RbcL C-terminal peptide is assumed to have a regulatory function in catalysis and cycle during the enzymatic reaction between a more open conformation and a tightly bound state, stabilizing the lid segment of the active site to enclose the substrate ribulose-bisphosphate (or the transition state analog, carboxyarabinitol bisphosphate [CABP]; (Duff et al., 2000; Zhu, 1998). Asp468 of RbcL, located adjacent to the RbcX recognition motif, participates in the formation of the closed state. In the crystal structure of Rubisco from *Synechococcus* sp. PCC6301 (in complex with CABP; Figure 50A), the C-terminal RbcL peptide is not covered by RbcS but would nevertheless be inaccessible to RbcX as it is not sufficiently detached from the complex, and its hydrophobic side chains are only partially solvent exposed. Thus, RbcX may interact with the C-terminal sequence until its attachment to the main body of Rubisco occurs (Saschenbrecker et al., 2007).



It was also demonstrated that RbcX failed to recognize the RbcL C-terminal peptide (EIKFEFD) with phenylalanines mutated to alanines (EIKAEAD) (Saschenbrecker et al., 2007). So the specific recognition of the C-terminal RbcL peptide by a central RbcX-binding cleft is critical for the function of RbcX. This C-terminal peptide sequence is highly conserved in organisms possessing Form I Rubisco and RbcX (Figure 43A). This sequence is absent in the RbcL of organisms possessing Form II Rubisco, e.g., *Rhodospirillum rubrum* (Figure 43C). Indeed, the translation of RbcL from *Synechococcus* sp. PCC6301 in the presence of Syn7002-RbcX resulted in the enhanced production of RbcL<sub>8</sub> core complexes (Figure 43B, Lane 5-6). Similarly, translation of RbcL from *Synechococcus* sp. PCC7002 resulted in the production of RbcL<sub>8</sub> complex only when RbcX was added to the lysate (Figure 43B, Lane 7-9). However, Syn7002-RbcX could not enhance the yield of *R. rubrum* dimer (Figure 43B, Lane 2-3), indicating that the RbcL C-terminal peptide sequence is necessary for interaction with RbcX and this interaction is a universal mechanism in the organisms containing Form I Rubisco and RbcX, such as cyanobacteria and higher plants.



**Figure 50. Crystal structure of Syn6301-RbcL<sub>8</sub>S<sub>8</sub> and Model of RbcL<sub>8</sub>-RbcX complex.**

A. Molecular representation of Rubisco complex of *Synechococcus* sp. PCC6301 bound to CABP (Newman et al., 1993). RbcL subunits are shown in surface representation, with the exception of the C-terminal peptides comprising residues 460-475 which is indicated as blue coils. This conformation is found only in

complex with the enzyme substrate. Subunits comprising RbcL dimers are colored white and blue; residues interacting with RbcS are shown in red. The small subunits themselves are shown as green ribbons.

B. Hypothetical model of binding of eight RbcX<sub>2</sub> to the assembled RbcL<sub>8</sub> core using similar representation as in A. the model is based on the structures of Syn6301-RbcL<sub>8</sub>S<sub>8</sub> and the RbcX-peptide complex. One edge of RbcX fills the gap between adjacent RbcL dimers, masking most of the surface covered by the globular domain of RbcS in the holoenzyme complex. In this arrangement, the C-termini of RbcL (shown in red) kink horizontally immediately N-terminal to the recognition motif, leaving the secondary structure composition of RbcL intact. (These figures were prepared by Dr. Andreas Bracher).

### 6.3 Implications for Rubisco assembly in higher plants

The C-terminal RbcX recognition sequence is conserved in higher plant RbcL (Figure 43), and homologs to *rbcX* genes are present in plant genomes exhibiting ~30% sequence similarity to the cyanobacterial RbcX (Figure 34 and 44A). This suggests a general role of RbcX in ensuring the assembly of RbcL<sub>8</sub> and in turn RbcL<sub>8</sub>S<sub>8</sub>.

The secondary structure elements for *A. thaliana* RbcX were similar to cyanobacterial RbcX (Figure 45). The functional studies revealed that *A. thaliana* RbcX can functionally replace cyanobacterial RbcX in assisting the assembly of cyanobacterial Rubisco, despite their low sequence identity (Figure 44A). However, the preliminary attempts to express *A. thaliana* Rubisco as an active enzyme in the *E.coli* based translation system *in vitro*, in the presence of RbcX and chaperonins, were not successful (Figure 48, Lanes 3-11), suggesting the need for an additional factor for plant Rubisco assembly.

The RbcX mechanism may serve as a model for explaining the efficient assembly of oligomeric protein complexes in the cell by specialized cellular machinery. It also explains the general route the assembly chaperones follow to perform their function by recognizing the surface exposed N- or C-terminal segments. It seems plausible that the function of RbcX needs to be considered when attempting to improve the catalytic properties of crop plant Rubisco by protein engineering.



## 7. References

- Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117, 199-209.
- Andersson, I. (2008). Catalysis and regulation in Rubisco. *Journal of experimental botany* 59, 1555-1568.
- Andersson, I., and Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiol Biochem* 46, 275-291.
- Andersson, I., and Taylor, T.C. (2003). Structural framework for catalysis and regulation in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of biochemistry and biophysics* 414, 130-140.
- Andrews, T.J. (1988). Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *The Journal of biological chemistry* 263, 12213-12219.
- Andrews, T.J., and Lorimer, G.H. (1985). Catalytic properties of a hybrid between cyanobacterial large subunits and higher plant small subunits of ribulose bisphosphate carboxylase-oxygenase. *The Journal of biological chemistry* 260, 4632-4636.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science (New York, NY)* 181, 223-230.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1997). *Current Protocols in Molecular Biology* (NY, John Wiley & Sons).
- Azem, A., Diamant, S., Kessel, M., Weiss, C., and Goloubinoff, P. (1995). The protein-folding activity of chaperonins correlates with the symmetric GroEL14(GroES7)2 heterooligomer. *Proceedings of the National Academy of Sciences of the United States of America* 92, 12021-12025.
- Azem, A., Kessel, M., and Goloubinoff, P. (1994). Characterization of a functional GroEL14(GroES7)2 chaperonin hetero-oligomer. *Science (New York, NY)* 265, 653-656.
- Baker, T.S., Eisenberg, D., and Eiserling, F. (1977). Ribulose Bisphosphate Carboxylase: A Two-Layered, Square-Shaped Molecule of Symmetry 422. *Science (New York, NY)* 196, 293-295.
- Baneyx, F., Bertsch, U., Kalbach, C.E., van der Vies, S.M., Soll, J., and Gatenby, A.A. (1995). Spinach chloroplast cpn21 co-chaperonin possesses two functional domains

## REFERENCES

---

fused together in a toroidal structure and exhibits nucleotide-dependent binding to plastid chaperonin 60. *The Journal of biological chemistry* 270, 10695-10702.

Beatrix, B., Sakai, H., and Wiedmann, M. (2000). The alpha and beta subunit of the nascent polypeptide-associated complex have distinct functions. *The Journal of biological chemistry* 275, 37838-37845.

Ben-Zvi, A.P., and Goloubinoff, P. (2001). Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *Journal of structural biology* 135, 84-93.

Bertsch, U., and Soll, J. (1995). Functional analysis of isolated cpn10 domains and conserved amino acid residues in spinach chloroplast co-chaperonin by site-directed mutagenesis. *Plant molecular biology* 29, 1039-1055.

Bertsch, U., Soll, J., Seetharam, R., and Viitanen, P.V. (1992). Identification, characterization, and DNA sequence of a functional "double" groES-like chaperonin from chloroplasts of higher plants. *Proceedings of the National Academy of Sciences of the United States of America* 89, 8696-8700.

Boston, R.S., Viitanen, P.V., and Vierling, E. (1996). Molecular chaperones and protein folding in plants. *Plant molecular biology* 32, 191-222.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72, 248-254.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., and Sigler, P.B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371, 578-586.

Brinker, A., Pfeifer, G., Kerner, M.J., Naylor, D.J., Hartl, F.U., and Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* 107, 223-233.

Bukau, B., Deuerling, E., Pfund, C., and Craig, E.A. (2000). Getting newly synthesized proteins into shape. *Cell* 101, 119-122.

Calvin, M. (1989). Forty years of photosynthesis and related activities. *Photosynth Res* 21, 3-16.

Catanzariti, A.M., Soboleva, T.A., Jans, D.A., Board, P.G., and Baker, R.T. (2004). An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci* 13, 1331-1339.

## REFERENCES

---

- Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *Journal of Biological Chemistry* 261, 12414-12419.
- Checa, S.K., and Viale, A.M. (1997). The 70-kda heat-shock protein dnaK chaperone system is required for the productive folding of ribulose-bisphosphate carboxylase subunits in *Escherichia coli*. *European Journal of Biochemistry* 248, 848-855.
- Cliff, M.J., Kad, N.M., Hay, N., Lund, P.A., Webb, M.R., Burston, S.G., and Clarke, A.R. (1999). A kinetic analysis of the nucleotide-induced allosteric transitions of GroEL. *J Mol Biol* 293, 667-684.
- Coligan, J.E., Dunn, B.M., Speicher, D.W., Wingfield, P.T., Ploegh, H.L. (2000). *Current Protocols in Protein Science* (NY, John Wiley & Sons).
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16, 10881-10890.
- Curmi, P.M., Cascio, D., Sweet, R.M., Eisenberg, D., and Schreuder, H. (1992). Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 2.0-Å resolution. *The Journal of biological chemistry* 267, 16980-16989.
- de la Fortelle, E., Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* 276, 472-494.
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* 400, 693-696.
- Dickson, R., Weiss, C., Howard, R.J., Alldrick, S.P., Ellis, R.J., Lorimer, G., Azem, A., and Viitanen, P.V. (2000). Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. *The Journal of biological chemistry* 275, 11829-11835.
- Duff, A.P., Andrews, T.J., and Curmi, P.M. (2000). The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J Mol Biol* 298, 903-916.
- Ecroyd, H., and Carver, J.A. (2008). Unraveling the mysteries of protein folding and misfolding. *IUBMB life* 60, 769-774.
- Ellis, R.J. (1979). The most abundant protein in the world. *Trends in biochemical sciences* 4, 241-244.

## REFERENCES

---

- Ellis, R.J. (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Current opinion in structural biology* 11, 114-119.
- Ellis, R.J. (2005). Chaperomics: in vivo GroEL function defined. *Curr Biol* 15, R661-663.
- Ellis, R.J. (2006). Molecular chaperones: assisting assembly in addition to folding. *Trends in biochemical sciences* 31, 395-401.
- Ellis, R.J., and van der Vies, S. M. (1991). Molecular chaperones. *Annu Rev Biochem* 60, 321-347.
- Ellis, R.J., and Hartl, F.U. (1996). Protein folding in the cell: competing models of chaperonin function. *Faseb J* 10, 20-26.
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 8, 978-984.
- Emlyn-Jones, D., Woodger, F.J., Price, G.D., and Whitney, S.M. (2006). RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant & cell physiology* 47, 1630-1640.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *Journal of molecular graphics & modelling* 15, 132-134, 112-133.
- Evans, P.R. (1997). Scala. *Joint CCP4 and ESF-EACBM Newsletter* 33, 22-24.
- Ewalt, K.L., Hendrick, J.P., Houry, W.A., and Hartl, F.U. (1997). In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* 90, 491-500.
- Finn, M.W., and Tabita, F.R. (2003). Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. *Journal of bacteriology* 185, 3049-3059.
- Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annual review of biochemistry* 70, 603-647.
- Gatenby, A.A., and Ellis, R.J. (1990). Chaperone function: the assembly of ribulose bisphosphate carboxylase-oxygenase. *Annual review of cell biology* 6, 125-149.
- Gatenby, A.A., Lubben, T.H., Ahlquist, P., and Keegstra, K. (1988). Imported large subunits of ribulose bisphosphate carboxylase/oxygenase, but not imported beta-ATP

## REFERENCES

---

synthase subunits, are assembled into holoenzyme in isolated chloroplasts. *The EMBO journal* 7, 1307-1314.

Gatenby, A.A., Viitanen, P.V., and Lorimer, G.H. (1990). Chaperonin assisted polypeptide folding and assembly: implications for the production of functional proteins in bacteria. *Trends in biotechnology* 8, 354-358.

Gautschi, M., Lilie, H., Funfschilling, U., Mun, A., Ross, S., Lithgow, T., Rucknagel, P., and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zutotin. *Proceedings of the National Academy of Sciences of the United States of America* 98, 3762-3767.

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P.S., Hartl, F.U., and Georgopoulos, C. (2004). In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Reports* 5, 195-200.

Georgopoulos, C. (1992). The emergence of the chaperone machines. *Trends in biochemical sciences* 17, 295-299.

Gething, M.J., and Sambrook, J. (1992). Protein folding in the cell. *Nature* 355, 33-45.

Gill, S.C., and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry* 182, 319-326.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A., and Lorimer, G.H. (1989a). Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* 342, 884-889.

Goloubinoff, P., Gatenby, A.A., and Lorimer, G.H. (1989b). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* 337, 44-47.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics (Oxford, England)* 15, 305-308.

Gutsche, I., Essen, L.O., and Baumeister, W. (1999). Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J Mol Biol* 293, 295-312.

Gutteridge, S., and Gatenby, A.A. (1995). Rubisco Synthesis, Assembly, Mechanism, and Regulation. *The Plant cell* 7, 809-819.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557-580.

Hansen, W.J., Cowan, N.J., and Welch, W.J. (1999). Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. *The Journal of cell biology* 145, 265-277.

## REFERENCES

---

- Hanson, T.E., and Tabita, F.R. (2001). A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4397-4402.
- Harris, J.R. (1991). Negative staining-carbon film technique: new cellular and molecular applications. *Journal of electron microscopy technique* 18, 269-276.
- Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. *Nature* 381, 571-579.
- Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science (New York, NY)* 295, 1852-1858.
- Hartman, F.C., and Harpel, M.R. (1994). Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annual review of biochemistry* 63, 197-234.
- Hegerl, R. (1996). The EM Program Package: A Platform for Image Processing in Biological Electron Microscopy. *Journal of structural biology* 116, 30-34.
- Hegerl, R., and Altbauer, A. (1982). The "EM" program system. *Ultramicroscopy* 9, 109-116.
- Hemmingsen, S.M., and Ellis, R.J. (1986). Purification and Properties of Ribulosebisphosphate Carboxylase Large Subunit Binding Protein. *Plant physiology* 80, 269-276.
- Hemmingsen, S.M., Woolford, C., van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W., and Ellis, R.J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 333, 330-334.
- Hesterkamp, T., Hauser, S., Lutcke, H., and Bukau, B. (1996). *Escherichia coli* trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America* 93, 4437-4441.
- Hill, J.E., and Hemmingsen, S.M. (2001). *Arabidopsis thaliana* type I and II chaperonins. *Cell stress & chaperones* 6, 190-200.
- Hirohashi, T., Nishio, K., and Nakai, M. (1999). cDNA sequence and overexpression of chloroplast chaperonin 21 from *Arabidopsis thaliana*. *Biochimica et biophysica acta* 1429, 512-515.
- Hohn, T., Hohn, B., Engel, A., Wurtz, M., and Smith, P.R. (1979). Isolation and characterization of the host protein groE involved in bacteriophage lambda assembly. *J Mol Biol* 129, 359-373.

## REFERENCES

---

- Houtz, R.L., and Portis, A.R., Jr. (2003). The life of ribulose 1,5-bisphosphate carboxylase/oxygenase--posttranslational facts and mysteries. *Archives of biochemistry and biophysics* 414, 150-158.
- Hubbs, A.E., and Roy, H. (1993). Assembly of invitro synthesized large subunits into ribulose-bisphosphate carboxylase oxygenase - formation and discharge of an I(8)-like species. *Journal of Biological Chemistry* 268, 13519-13525.
- Ishii, N., Taguchi, H., Sumi, M., and Yoshida, M. (1992). Structure of holo-chaperonin studied with electron microscopy. Oligomeric cpn10 on top of two layers of cpn60 rings with two stripes each. *FEBS letters* 299, 169-174.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta crystallographica* 47 ( Pt 2), 110-119.
- Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., et al. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122, 209-220.
- Kitano, K., Maeda, N., Fukui, T., Atomi, H., Imanaka, T., and Miki, K. (2001). Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. *Structure* 9, 473-481.
- Kleywegt, G.T., Jones, T.A. (1994). A super position. *CCP4/ESF-EACBM Newsletter on Protein Crystallography* 31, 9-14.
- Knight, S., Andersson, I., and Branden, C.I. (1990). Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. Subunit interactions and active site. *J Mol Biol* 215, 113-160.
- Koumoto, Y., Shimada, T., Kondo, M., Hara-Nishimura, I., and Nishimura, M. (2001). Chloroplasts have a novel Cpn10 in addition to Cpn20 as co-chaperonins in *Arabidopsis thaliana*. *The Journal of biological chemistry* 276, 29688-29694.
- Koumoto, Y., Shimada, T., Kondo, M., Takao, T., Shimonishi, Y., Hara-Nishimura, I., and Nishimura, M. (1999). Chloroplast Cpn20 forms a tetrameric structure in *Arabidopsis thaliana*. *Plant J* 17, 467-477.
- Kraulis, P.J. (1991). MOLSCRIPT - A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24, 946-950.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.



## REFERENCES

---

- Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., and Hartl, F.U. (1992a). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* 356, 683-689.
- Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F.U. (1992b). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *The EMBO journal* 11, 4757-4765.
- Lanzetta, P.A., Alvarez, L.J., Reinach, P.S., and Candia, O.A. (1979). An improved assay for nanomole amounts of inorganic phosphate. *Analytical biochemistry* 100, 95-97.
- Larimer, F.W., and Soper, T.S. (1993). Overproduction of *Anabaena* 7120 ribulose-bisphosphate carboxylase/oxygenase in *Escherichia coli*. *Gene* 126, 85-92.
- Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275, 416-420.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26 283-291.
- Lee, B., and Tabita, F.R. (1990). Purification of recombinant ribulose-1,5-bisphosphate carboxylase/oxygenase large subunits suitable for reconstitution and assembly of active  $L_8S_8$  enzyme. *Biochemistry* 29, 9352-9357.
- Leroux, M.R., and Hartl, F.U. (2000). Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr Biol* 10, R260-264.
- Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletter* 26.
- Levinthal, C. (1968). Are there pathways for protein folding? *J Chim Phys* 65, 44-45.
- Li, H., Sawaya, M.R., Tabita, F.R., and Eisenberg, D. (2005). Crystal structure of a RuBisCO-like protein from the green sulfur bacterium *Chlorobium tepidum*. *Structure* 13, 779-789.
- Li, L.A., and Tabita, F.R. (1997). Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *Journal of bacteriology* 179, 3793-3796.
- Li, R., and Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci* 8, 1571-1590.

## REFERENCES

---

- Lilley, R.M., and Portis, A.R. (1990). Activation of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) by Rubisco Activase : Effects of Some Sugar Phosphates. *Plant physiology* 94, 245-250.
- Maeda, N., Kitano, K., Fukui, T., Ezaki, S., Atomi, H., Miki, K., and Imanaka, T. (1999). Ribulose bisphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. *J Mol Biol* 293, 57-66.
- Martel, R., Cloney, L.P., Pelcher, L.E., and Hemmingsen, S.M. (1990). Unique composition of plastid chaperonin-60: alpha and beta polypeptide-encoding genes are highly divergent. *Gene* 94, 181-187.
- Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L., and Hartl, F.U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. *Nature* 352, 36-42.
- Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* 62, 670-684.
- Meyer, A.S., Gillespie, J.R., Walther, D., Millet, I.S., Doniach, S., and Frydman, J. (2003). Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* 113, 369-381.
- Milos, P., and Roy, H. (1984). ATP-released large subunits participate in the assembly of RuBP carboxylase. *J Cell Biochem* 24, 153-162.
- Murphy, R.M. (2002). Peptide aggregation in neurodegenerative disease. *Annual review of biomedical engineering* 4, 155-174.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.
- Musgrove, J.E., Johnson, R.A., and Ellis, R.J. (1987). Dissociation of the ribulosebisphosphate-carboxylase large-subunit binding protein into dissimilar subunits. *European journal of biochemistry / FEBS* 163, 529-534.
- Newman, J., and Gutteridge, S. (1993). The X-ray structure of *Synechococcus* ribulose-bisphosphate carboxylase/oxygenase-activated quaternary complex at 2.2-A resolution. *The Journal of biological chemistry* 268, 25876-25886.
- Onizuka, T., Endo, S., Akiyama, H., Kanai, S., Hirano, M., Yokota, A., Tanaka, S., and Miyasaka, H. (2004). The *rbcX* gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant & cell physiology* 45, 1390-1395.

## REFERENCES

---

- Ort, D.R. (1994). Photosynthesis. *Encyclopedia of Agricultural Sciences* 3 187-195.
- Patzelt, H., Rudiger, S., Brehmer, D., Kramer, G., Vorderwulbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J., et al. (2001). Binding specificity of *Escherichia coli* trigger factor. *Proceedings of the National Academy of Sciences of the United States of America* 98, 14244-14249.
- Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nature structural biology* 6, 458-463.
- Portis, A.R., Jr. (1990). Rubisco activase. *Biochimica et biophysica acta* 1015, 15-28.
- Ranson, N.A., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L., and Saibil, H.R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107, 869-879.
- Roessner, D., Kulicke, W.M. (1994). On-line coupling of flow field-flow fractionation and multiangle laser light scattering. *J Chromatogr A*. 687, 249-258.
- Roseman, A.M., Ranson, N.A., Gowen, B., Fuller, S.D., and Saibil, H.R. (2001). Structures of unliganded and ATP-bound states of the *Escherichia coli* chaperonin GroEL by cryoelectron microscopy. *Journal of structural biology* 135, 115-125.
- Rudi, K., Skulberg, O.M., and Jakobsen, K.S. (1998). Evolution of cyanobacteria by exchange of genetic material among phyletically related strains. *Journal of bacteriology* 180, 3453-3461.
- Rudiger, S., Schneider-Mergener, J., and Bukau, B. (2001). Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *The EMBO journal* 20, 1042-1050.
- Saibil, H., Dong, Z., Wood, S., and auf der Mauer, A. (1991). Binding of chaperonins. *Nature* 353, 25-26.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning*. Cold Spring Harbor, NY
- Schagger, H., and von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Analytical biochemistry* 199, 223-231.
- Schmidt, M., Buchner, J., Todd, M. J., Lorimer, G. H., and Viitanen, P. V. (1994). On the role of groES in the chaperonin-assisted folding reaction. Three case studies. *J Biol Chem* 269, 10304-10311.

## REFERENCES

---

- Schneider, G., Knight, S., Andersson, I., Branden, C.I., Lindqvist, Y., and Lundqvist, T. (1990). Comparison of the crystal structures of  $L_2$  and  $L_8S_8$  Rubisco suggests a functional role for the small subunit. *The EMBO journal* 9, 2045-2050.
- Schneider, G., Lindqvist, Y., and Branden, C.I. (1992). RUBISCO: structure and mechanism. *Annual review of biophysics and biomolecular structure* 21, 119-143.
- Schultz, C.P. (2000). Illuminating folding intermediates. *Nature structural biology* 7, 7-10.
- Sharkia, R., Bonshtien, A.L., Mizrahi, I., Weiss, C., Niv, A., Lustig, A., Viitanen, P.V., and Azem, A. (2003). On the oligomeric state of chloroplast chaperonin 10 and chaperonin 20. *Biochimica et biophysica acta* 1651, 76-84.
- Siegers, K., Waldmann, T., Leroux, M.R., Grein, K., Shevchenko, A., Schiebel, E., and Hartl, F.U. (1999). Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system. *The EMBO journal* 18, 75-84.
- Somerville, C.R., McIntosh, L., Fitchen, J., Gurevitz, M. (1986). The cloning and expression in *Escherichia coli* of RuBP carboxylase/oxygenase large subunit genes. *Methods Enzymol* 118, 419-433.
- Spiess, C., Meyer, A.S., Reissmann, S., and Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends in cell biology* 14, 598-604.
- Spreitzer, R.J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of biochemistry and biophysics* 414, 141-149.
- Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res* 60, 1-28.
- Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56, 965-972.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55, 849-861.
- Tilly, K., Murialdo, H., and Georgopoulos, C. (1981). Identification of a second *Escherichia coli* groE gene whose product is necessary for bacteriophage morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 78, 1629-1633.
- Tsai, C.J., Ma, B., and Nussinov, R. (1999). Folding and binding cascades: shifts in energy landscapes. *Proceedings of the National Academy of Sciences of the United States of America* 96, 9970-9972.

## REFERENCES

---

- Tsuprun, V.L., Boekema, E.J., Samsonidze, T.G., and Pushkin, A.V. (1991). Electron microscopy of the complexes of ribulose-1,5-bisphosphate carboxylase (Rubisco) and Rubisco subunit-binding protein from pea leaves. *FEBS letters* 289, 205-209.
- Vagin, A.A., and Isupov, M.N. (2001). Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* 57, 1451-1456.
- Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L., and Cowan, N.J. (1998). Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863-873.
- Valpuesta, J.M., Martin-Benito, J., Gomez-Puertas, P., Carrascosa, J.L., and Willison, K.R. (2002). Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. *FEBS letters* 529, 11-16.
- Vendruscolo, M., and Dobson, C.M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. *Philosophical transactions* 363, 433-450; discussion 450-432.
- Viitanen, P.V., Schmidt, M., Buchner, J., Suzuki, T., Vierling, E., Dickson, R., Lorimer, G.H., Gatenby, A., and Soll, J. (1995). Functional characterization of the higher plant chloroplast chaperonins. *The Journal of biological chemistry* 270, 18158-18164.
- Walter, S., and Buchner, J. (2002). Molecular chaperones--cellular machines for protein folding. *Angewandte Chemie (International ed)* 41, 1098-1113.
- Watson, G.M., Yu, J.P., and Tabita, F.R. (1999). Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea. *Journal of bacteriology* 181, 1569-1575.
- Weiner, M.P., Costa, G.L., Schoettlin, W., Cline, J., Mathur, E., and Bauer, J.C. (1994). Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* 151, 119-123.
- Weissman, J.S., Hohl, C.M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H.R., Fenton, W.A., and Horwich, A.L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* 83, 577-587.
- Wiedmann, B., Sakai, H., Davis, T.A., and Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. *Nature* 370, 434-440.
- Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. *Anal Chim Acta* 272, 1-40.
- Yang, T., and Poovaiah, B.W. (2000). Arabidopsis chloroplast chaperonin 10 is a calmodulin-binding protein. *Biochemical and biophysical research communications* 275, 601-607.

## REFERENCES

---

Yoon, M., Putterill, J.J., Ross, G.S., and Laing, W.A. (2001). Determination of the relative expression levels of rubisco small subunit genes in Arabidopsis by rapid amplification of cDNA ends. *Analytical biochemistry* 291, 237-244.

Young, J.C., Barral, J.M., and Hartl, F.U. (2003). More than folding: localized functions of cytosolic chaperones. *Trends in biochemical sciences* 28, 541-547.

Zhu, G.H., Jensen, R.G., Bohnert, H.J., Wildner, G.F., and Schlitter, J. (1998). Dependence of catalysis and CO<sub>2</sub>/O<sub>2</sub> specificity of Rubisco on the carboxy-terminus of the large subunit at different temperatures. *Photosynth Res* 57, 71-79.

## 8. APPENDICES

### 8.1 Abbreviations

AA	Acrylamide
aa	amino acid
ADP	Adenosine 5'-diphosphate
Amp	Ampicillin
Amp <sup>R</sup>	Ampicillin resistance
AMP-PNP	Adenosine 5'-( $\beta,\gamma$ -imido)triphosphate
Ana	<i>Anabaena</i>
ANS	1-anilino-8-naphthalene-sulphonate
APS	Ammonium persulfate
At	<i>Arabidopsis thaliana</i>
ATP	Adenosine 5'-triphosphate
AU	Arbitrary unit
BLAST	Basic Local Alignment Search Tool
$\beta$ -NADH	$\beta$ -nicotinamide adenine dinucleotide
bp	base pairs
BSA	Bovine serum albumin
$^{\circ}\text{C}$	degree Celsius
$^{14}\text{C}$	Carbon-14
Cam	Chloramphenicol
Cam <sup>R</sup>	Chloramphenicol resistance
CAPB	Carboxy-arabinitol-1,5-bisphosphate
cDNA	copy DNA
CDTA	trans-1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid
Ch	Chloroplast
CIP	Calf intestinal alkaline phosphatase
Cpn	Chaperonin
cryo-EM	Cryoelectron microscopy
$\Delta$ (delta)	deletion
DE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
DnaJ	Bacterial Hsp40 chaperone
DnaK	Bacterial Hsp70 chaperone
dNTP	2'-desoxyribonucleotide-triphosphate
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence



## APPENDICES

---

E. coli	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
FFF	Field flow fractionation
Fig.	Figure
FPLC	Fast protein liquid chromatography
FT	flowthrough
g	Acceleration of gravity, 9.81 m/s <sup>2</sup>
GdnHCl	Guanidinium hydrochloride
GroEL	Bacterial Hsp60 chaperonin
GroES	Bacterial Hsp10 co-chaperone
GrpE	Bacterial nucleotide exchange factor of DnaK
GSH	reduced glutathione
GSSG	oxidized glutathione
h	hour
HEPES	N-(2-hydroxyethyl)piperazin-N'-2-ethanesulfonic acid
His6	Hexa histidine-tag
HRP	Horseradish peroxidase
Hsp	Heat shock protein
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
Kan <sup>R</sup>	Kanamycin resistance
KD	Dissociation constant
kDa	kilodalton
LB	Luria Bertani
MDH	Malate dehydrogenase
MES	2-morpholinoethanesulfonic acid
Met	Methionine
MLS	Multiangle light scattering
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
mt	mitochondrial
MW	Molecular weight
MWCO	Molecular weight cut off
N-	N-terminal
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate
Ni-NTA	Nickel-nitrilotriacetic acid
OAc	acetate
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis

## APPENDICES

---

PCR	Polymerase chain reaction
PEG	Polyethylene glycol
Pfu	<i>Pyrococcus furiosus</i>
pH	reverse logarithm of relative hydrogen proton (H <sup>+</sup> ) concentration
Pi	inorganic phosphate
PK	Proteinase K
PMSF	Phenyl-methyl-sulfonyl fluoride
RbcL	RuBisCO large subunit
RbcS	RuBisCO small subunit
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
PVDF	Polyvinylidenfluorid
RNA	Ribonucleic acid
RNase A	Ribonuclease A
rpm	revolutions per minute
Rr	<i>Rhodospirillum rubrum</i>
RT	room temperature
RTS	Rapid Transcription/Translation System
<sup>35</sup> S	Sulphur-35
SDS	Sodium dodecyl sulfate
SeMet	selenomethionine
So	<i>Spinacia oleracea</i>
sp.	species
Syn	<i>Synechococcus</i>
TAE	Tris-acetate-EDTA (EDTA; Tris)
TBS	Tris-buffered saline (Tris)
TBS-T	TBS containing Tween20 (TBS; Tween20)
TCA	Trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
TEV	Tobacco etch virus
TF	Trigger factor
Tris	Trishydroxymethylaminomethane
TritonX-100	octyl phenol ethoxylate
Tween20	polyoxyethylene-sorbitan-monolaurate
tRNA	transfer-RNA
UV/VIS	Ultraviolet/visible
w/o	without
v/v	volume per volume
w/v	weight per volume

### 8.3 Publications

Saschenbrecker, S., Bracher, A., Rao, K.V., **Rao, B.V.**, Hartl, F.U., and Hayer-Hartl, M. (2007). Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* 129, 1189-1200.

Behrends, C., Langer, C.A., Boteva, R., Bottcher, U.M., Stemp, M.J., Schaffar, G., **Rao, B.V.**, Giese, A., Kretschmar, H., Siegers, K., et al. (2006). Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Molecular cell* 23, 887-897.

Das, M., **Rao, B.V.**, Ghosh, S., and Varadarajan, R. (2005). Attempts to delineate the relative contributions of changes in hydrophobicity and packing to changes in stability of ribonuclease S mutants. *Biochemistry* 44, 5923-5930.

## 8.4 Curriculum Vitae

Name: Bharathi Vasudeva Rao  
Date of Birth: 26 January 1979  
Birthplace: Kolar, Karnataka, India  
Nationality: Indian  
Gender: Female  
Marital Status: Married  
Spouse name: Karnam Vasudeva Rao

### Academic Details:

1996-1999	Bachelors in science (B. Sc.) in Chemistry, Botany and Zoology, Bangalore University, Bangalore, India.
1999-2001	Master of Science (M. Sc.) in Botany with specialization in Microbiology, St. Joseph's college, Bangalore University, Bangalore, India.
Oct 2001- May 2003	Worked as Project assistant under Dr. Raghavan Varadarajan, Molecular Biophysics unit, Indian Institute of science, Bangalore, India. Involved in research work on studying the importance of the hydrophobic effect vs. packing interactions in protein folding.
July 2003 - Oct 2003	Worked as Research Assistant under Dr. Ming Tsair Chan, Institute of Bioagricultural sciences, Academia Sinica, Taipei, Taiwan (R.O.C.). Involved in the research work on cloning and transformation of <i>ICE1</i> (stress related gene from <i>Arabidopsis</i> ) in Jute ( <i>Corchorus</i> spp.).
Aug 2004 - present date	PhD research under Prof. Ulrich Hartl, Department of Cellular Biochemistry, Max Planck Institute for Biochemistry, Martinsried, Munich. Involved in research on structural characterization of chaperone assisted RuBisCO folding and assembly.